Investigating facial shape using Multilevel Principal Component Analysis

A thesis presented in fulfilment for the requirements of the degree of Doctor of Philosophy (PhD)

Jennifer Galloway

BDS, MDSc, BMSc (Hons), MFDS RCPS (Glasg), MOrth RCS (Edin)

Cardiff University School of Dentistry

2021

DECLARATION AND STATEMENTS

Statement 1: This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed _____ Date _____

Statement 2: This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation).

Signed _____ Date _____

Statement 3: I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

Signed _____ Date _____

Declaration: This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third-Party Editors by Research Degree Students Procedure.

Signed _____ Date _____

Word Count: 42940 (Excluding summary, acknowledgements, declarations, contents pages, appendices, tables, diagrams and figures, references, bibliography, footnotes and endnotes)

ACKNOWLEDGEMENTS

I am extremely grateful to all the families who took part in the ALSPAC, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

I would also like to thank my supervisors (Professor Richmond, Dr Farnell and Dr Zhurov) and all those at Cardiff University for facilitating the completion of this PhD during my clinical training. The additional guidance of the team in KU Leuven has been invaluable.

I would also like to thank my family for their continual support, patience and guidance.

SUMMARY

Aims:

1. To determine the influence of geographical location, sex, height, Body Mass Index (BMI), age (14-16 years old), pubertal stage, metabolic factors, atopy, breathing disorders, maternal smoking and alcohol consumption during pregnancy on facial shape.
2. To explore the usefulness of Multilevel Principal Component Analysis (mPCA) in facial shape research.

Method: The influence of geographical location and sex was assessed using 21 landmarks on 3D facial scans of subjects from Croatia (n=73), England (n=79), Wales (n=50) and Finland (n=47). The influence of sex, height, BMI, age (14-16 years old), pubertal stage, metabolic factors, atopy, breathing disorders, maternal smoking and alcohol consumption during pregnancy on adolescent facial shape was assessed using 1000 and 7160 quasi-landmarks on 3D facial scans of the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort (n=1411). The results of mPCA were compared to those using landmarks only, conventional Principal Component Analysis (PCA), Discriminate Function Analysis (DFA) and Partial Least Squares Regression (PLSR). mPCA was also assessed as a variable selection tool prior to PLSR.

Results: mPCA provided more meaningful information in the exploratory phase of data analysis than conventional PCA and DFA. However, the results must be interpreted with caution when group sizes are imbalanced. All variables reached significance, except for age, in their respective mPCA models. Geographical location, sex, height, BMI and fasting insulin explained greater than 5% of the total variation. These variables also reached significance in the PLSR models. Therefore 5% may be a useful threshold for PLSR variable selection.

Conclusions: Sex, geographical location, height, BMI and fasting insulin had the most influence on facial shape. mPCA appears to be a useful tool for visualising the maximum variation between groups of subjects when group sizes are balanced and as a variable selection tool to inform more sophisticated models such as PLSR.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BMI	Body mass index
mPCA	Multilevel principal component analysis
ALSPAC	Avon Longitudinal Study of Parents and Children
PCA	Principal component analysis
DFA	Discriminate function analysis
PLSR	Partial least squares regression
3D	Three dimensional
MANOVA	Multivariate analysis of variance
IHH	Indian hedgehog
FGFR3	Fibroblast growth factor receptor 3
GWAS	Genome-wide association studies
SNPs	Single nucleotide polymorphisms
GIANT	Genetic Investigation of Anthropometric Traits
GH	Growth hormone
SHBG	Sex hormone-binding globulin
IGF-1	Insulin-like growth factor 1
GHBP	Growth hormone binding protein
TMJ	Temporomandibular joint
HDL	High density lipids
LDL	Low density lipids
PCs	Principal components
2D	Two dimensional
LADA	Latent autoimmune diabetes in adults
CL/P	Cleft lip with or without cleft palate
OMENS	Orbit, mandible, ear, nerve and soft tissue
USA	United States of America
FAS	Foetal alcohol syndrome
CT	Computed tomography
CBCT	Cone beam computed tomography
MRI	Magnetic resonance imaging
ANOVA	Analysis of variance
CVA	Canonical variates analysis
LDA	Linear discriminate analysis
QDA	Quadratic discriminate analysis
UK	United Kingdom
GPA	Generalised Procrustes analysis
ELISA	Enzyme linked immunosorbent assay

VLDL	Very low density lipids
DF	Discriminate function
SMOTE	Synthetic minority oversampling technique
ADASYN	Adaptive synthetic sampling
t-SNA	t-distributed stochastic neighbor embedding
SVM	Support vector machines
BRIM	Bootstrapped response imputation modelling
CAH	Congenital adrenal hyperplasia

1 INTRODUCTION

1.1 BACKGROUND

The development of facial shape is influenced by complex interactions between genetics and the environment. Within orthodontics, faces can be broadly categorised into Class I, II and III skeletal patterns, with rudimentary descriptions of facial height proportions, facial asymmetry and soft tissue features (Cobourne and DiBiase 2016). Alternative classification methods are summarised by Franco et al. (2013) which include: bradyfacial, mesofacial and dolichofacial; hypodivergent, neutral and hyperdivergent; and leptoprosopic, mesoprosopic and euryprosopic. Collett and West (1993) advise that the historical focus in orthodontics has been on the sagittal plane and suggest that others have been more concerned about the ratio between facial height and width by, for example, categorising into broad, medium and narrow using Cole's facial indices. Other fields describe the overall facial shape as heart, square, rectangular, round and oval shaped (Sunhem and Pasupa 2016). However, facial structures are far more complex and individual than broad categories suggest.

Previously, radiographs were used to investigate the craniofacial region, but this introduces the risk of radiation (Whaites 2002b). The advent of three-dimensional (3D) scanning has allowed for non-invasive research to be carried out (Toma et al. 2009). One of the challenges with 3D facial shape research is the large number of landmarks (outcome variables) that require analysing. This can range from a small number of landmarks (e.g., 21 landmarks) to tens of thousands of landmarks which provide a detailed analysis of surface facial shape. However, once the number of landmarks exceeds the number of subjects available for analysis and due to the correlated nature of the landmark data, traditional statistical techniques such as multivariate analysis of variance (MANOVA) and multiple regression can lose their power (Tabachnick and Fidell 1996; Shrimpton et al. 2014). Researchers have subsequently utilised mathematical techniques that allow the number of outcome variables to be reduced. This is known as dimension reduction.

Techniques that have been used for dimension reduction include Principal Component Analysis (PCA), Discriminate Function Analysis (DFA) and Partial Least Squares

Regression (PLSR) (Fisher 1938; Wold 1966; Jolliffe 2002). These allow statistical analysis to be more manageable whilst minimising the amount of information lost during the dimension reduction process. Conventional PCA has a distinct disadvantage in that it does not consider any groupings during the dimension reduction of the landmark data (Farnell et al. 2017). Conventional PCA calculates Principal Components (PCs), which are linear representations of the landmark points describing the direction and magnitude of maximum variance in multiple dimensions (Jolliffe 2002). However, as the influence of all potential groupings is not considered during the dimension reduction process, it is difficult, if not impossible, to disentangle the overall influence of variables on facial shape. Furthermore, when the PCs that explain the least amount of variation are discarded, there is the potential to lose information. There is therefore the potential to reduce the amount of useful information available after PCA and it is more difficult to interpret the influence of variables on facial shape. By contrast, alternative techniques such as PLSR and DFA do take variables into account.

A further alternative method is Multilevel Principal Component Analysis (mPCA). The benefits of multilevel models have previously been suggested by Lecron et al. (2012) and Timmerman (2006). mPCA was subsequently developed by Farnell et al. (2016) for the analysis of dental radiographs and smiles. Initial work was then focused on discriminating between the facial shape of populations (Farnell et al. 2017). As mPCA takes variables into account within the dimension reduction process, it may allow more useful information to be retained and facilitate visualisations that allow for a succinct summary of the differences in facial shape compared to conventional PCA.

This thesis will explore the advantages and disadvantages of mPCA, within the context of surface face shape research, by comparing mPCA to the analyses possible with the landmarks alone, conventional PCA, DFA and PLSR. The influence of geographical location, sex, height, BMI, pubertal stage, metabolic factors, atopy, breathing disorders, maternal smoking during pregnancy and alcohol consumption during pregnancy on facial shape will be explored. It is hoped that this research will highlight the relative advantages and disadvantages of using mPCA for analysing facial shape from a clinician's perspective, whilst providing further information on the influence

of the above variables on facial development. Ultimately, this information could be used to improve more sophisticated models aimed at progressing orthodontic diagnoses, predicting potential success rates of treatment modalities, growth prediction, personalised orthodontics, improved forensic analyses and facial recognition.

1.2 OVERVIEW OF THESIS

This thesis begins by discussing the development of the facial shape. Previous growth theories are discussed alongside more recent findings on the genetic and environmental influences on facial shape. Methods for investigating the influence of these factors are presented, including imaging modalities and possible analytical techniques. The benefits of dimension reduction techniques are highlighted. This introduces the concept of mPCA, as well as more traditional techniques, including conventional PCA, DFA and PLSR.

The thesis is separated into three studies. The first compares the facial shape of four European groupings (Croatian, English, Welsh and Finnish) with respect to geographical location, sex and within-group variation (i.e., every other source of variation excluding geographical location and sex). Twenty-one facial landmarks were used in the analyses which provides an opportunity to demonstrate the advantages of mPCA for exploring categorical variables in comparison to the raw landmarks, conventional PCA and DFA. The second study focuses on the English population in more detail. mPCA is used to determine the relative importance of sex, age (14-16 years old), pubertal stage, height, BMI, metabolic factors, atopy, breathing disorders, maternal smoking during pregnancy and maternal alcohol consumption during pregnancy. The advantage of using mPCA to give an indication of the relative importance variables is highlighted. The issues surrounding imbalanced group sample sizes are also discussed. The third and final study introduces mPCA as an aid for variable selection prior to further analyses. PLSR is used to explore the variables that explained greater than 2% of the total variation in their respective mPCA models.

The facial shape differences explained by each variable are subsequently discussed in relation to the information available in the published literature. Suggestions are made

for improvements to the methodology to facilitate more robust investigation into the influence of the variables in the future. Recommendations are also made with regards to mPCA, including further investigation into imbalanced data sets and principal component retention.

1.3 NOVEL CONTRIBUTIONS

Prior to the start of this body of work, mPCA had only been used in a two-level capacity in facial shape research to assess between- and within-group variation. This thesis introduces a three-level mPCA model and compares mPCA to DFA and the results possible with the landmarks only for the first time. This thesis provides a novel use for mPCA to aid variable selection prior to further analyses and compares the relative importance of numerous variables on facial shape for the first time.

Contributions are also made with regards to clinical questions. Prior to this thesis, the author is only aware of one paper describing the influence of maternal smoking on non-syndromic facial variation which used PLSR to assess the facial shape of one year old children. The results were submitted as supplementary material (Muggli et al. 2017). The long-term influence of maternal smoking during pregnancy on non-syndromic facial shape has not previously been assessed and is explored here for the first time using adolescent facial shapes. The influence of pubertal stage on facial shape appears to be explored in detail for the first time in this thesis. The influence of metabolic factors on facial shape has, to the author's knowledge, only been explored in one previous paper which used conventional PCA and multiple regression (Djordjevic et al. 2013b). The influence of fasting insulin was greater than any other metabolic factor and given its association with growth hormone, is worthy of further investigation here.

1.4 RELATED PUBLICATIONS

The following have been published in peer-reviewed journals and relate directly to topics explored in this thesis:

- **Galloway, J.**, Farnell, D.J.J., Zhurov, A.I., Richmond, S. 2020. Multilevel analysis of the influence of maternal smoking and alcohol consumption on the

facial shape of English adolescents. *Journal of Imaging* 6(34), p.10.3390/jimaging6050034. https://doi.org/10.3390/jimaging6050034

- Farnell, D.J.J., **Galloway, J.,** Zhurov, A., Richmond, S., Perttiniemi, P., Katic, V. 2017. Initial results of multilevel principal components analysis of facial shape. *Medical Image and Understanding Analysis*. In: Valdes-Hernandez, M., Gonzalez-Castro, V. eds *Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science*, vol 723. Cham: Springer. https://doi.org/10.1007/978-3-319-60964-5_59

The following book chapter is also related to work contained in this thesis:

- Richmond, S., Wilson-Nagrani, C., Zhurov, A., Farnell, D.J.J., **Galloway, J.**, Ali, A.S., Pirttiniemi, P., Katic, V. 2018. Factors influencing facial shape. In: Huang, G.J., Richmond, S., Vig, K.W. eds *Evidence-Based Orthodontics*, New Jersey: John Wiley and Sons.

The following utilise mPCA and have been published in peer-reviewed journals however investigate topics out-with the scope of this thesis:

- Farnell, D.J.J., Richmond, S., **Galloway, J.**, Zhurov, A.I., Pirttiniemi, P., Heikkinen, T., Harila, V., Matthews, H., Claes, P. 2021. An exploration of adolescent facial shape changes with age via multilevel Partial Least Squares Regression. *Computer Methods and Programs in Biomedicine* 200, p. 105935. https://doi.org/10.1016/j.cmpb.2021.105935
- Farnell, D.J.J., Richmond, S., **Galloway, J.**, Zhurov, A.I., Pirttiniemi, P., Heikkinen, T., Harila, V., Matthews, H., Claes, P. 2020. Multilevel principal components analysis of three-dimensional facial growth in adolescents. *Computer Methods and Programs in Biomedicine* 133, p. 105272. https://doi.org/10.1016/j.cmpb.2019.105272
- Farnell, D.J.J., **Galloway, J.**, Zhurov, A.I., Richmond, S. 2020. Multilevel models of age-related changes in facial shape in adolescents. In: Zheng, Y., Williams, B., Chen, K. eds *Medical Image Understanding and Analysis. MIUA 2019. Communications in Computer and Information Science*, vol 1065. Cham: Springer. https://doi.org/10.1007/978-3-319-95921-4_18

- Farnell, D.J.J., **Galloway, J.,** Zhurov, A.I., Richmond, S., Marshall, D., Rosin, P.L., Al Meyah, K., Pirttiniemi, P., Lahdesmaki, R. 2019. What's in a smile? Initial analyses of dynamic changes in facial shape and appearance. *Journal of Imaging* 5(1), p. 2. https://doi.org/10.3390/jimaging5010002
- Farnell, D.J.J., **Galloway, J.**, Zhurov, A., Richmond, S., Pirttiniemi, P., Lahdesmaki, R. 2018. What's in a smile? Initial results of multilevel principal components analysis of facial shape and image texture. In: Nixon, M., Mahmoodi, S., Zwiggelaar, R. eds. *Medical Image Understanding and Analysis. MIUA 2018. Communications in Computer and Information Science*, vol 894. Cham: Springer. https://doi.org/10.1007/978-3-319-95921-4_18

2 LITERATURE REVIEW

2.1 PRE-NATAL DEVELOPMENT OF THE CRANIOFACIAL REGION

The face begins development in utero, with two-thirds of the skull at its final dimension at birth (Medawar and Fisher 1944). It is therefore valuable to begin understanding facial development from the embryonic stage to determine all the factors that may be involved in the differences in facial shape between individuals. Embryos are derived from three germ layers: ectoderm, endoderm and mesoderm (Sadler 2012). However, it is the development of cranial neural crest cells during the formation of the neural tube, which allows for the development of important cranial structures (Sadler 2012). The prechordal plate subsequently develops as an area of thick endoderm in the region of the forebrain. This produces complex molecular signalling pathways which are initiated by homeobox genes. This allows for the development of two hemispheres in the brain, alongside many other important processes (Mercier et al. 2011).

The development of the face begins with the formation of pharyngeal arches during week four in utero (Sadler 2012). Each of these arches is composed of a band of mesenchyme with the surface facing the outside of the embryo composed of ectoderm and the surface facing the inside of the embryo composed of endoderm. The arches are joined by the ectodermal surface running continuously on the outer edge, producing clefts between the arches, and the endoderm following the same pattern on the inside of the embryo, producing pouches (Sadler 2012). A cartilage bar, aortic arch and cranial nerves run through the middle of each arch. Each arch develops distinct skeletal, muscular, vascular and soft tissue structures of the head and neck (Sadler 2012).

Mossey et al. (2009) describe the development of prominences from these arches during the sixth week in utero, which begin to form the facial features around the stomodeum (early mouth). The frontonasal prominence forms in the midline and becomes the forehead following signalling from the developing forebrain. Two nasal placodes subsequently develop which become nasal pits, and on either side of these placodes, medial and lateral nasal processes develop bilaterally. The lateral nasal

processes form the alae of the nose whilst the medial nasal processes develop into the tip of nose, philtrum and primary palate. These fuse with maxillary processes on either side, to complete the upper cheeks, upper lip and secondary palate. The face is completed by the fusion of two mandibular processes, which form the lower cheeks and mandible.

The bones of the cranium develop through mesenchyme condensation during the fourth-fifth week in utero and are ossified through intramembranous ossification (Jin et al. 2016). The cranial base ossifies through endochondral ossification, which begins with the development of primary cartilage at week six (Nie 2005). At birth, the face is smaller than would be expected in relation to the size of the skull due to underdevelopment of the nose and jaws. The processes that occur to this region after birth will therefore strongly influence facial shape at each stage of life.

2.2 POST-NATAL GROWTH OF THE CRANIOFACIAL REGION

The skull can be divided into the neurocranium and viscerocranium. The neurocranium includes the structures surrounding the brain and the viscerocranium primarily includes the face (Moore and Dalley 2006). Both have an influence on facial shape. The neurocranium expands with the increasing size of the brain. Bone is deposited at the sutures as they are passively displaced, whilst remodelling of the bony surface supplements this. The length of the cranial base increases through remodelling and endochondral growth (Melsen 1974) and will affect the shape of the face due to its relationship with the jaws and frontal bone. Most of the growth of the cranial base occurs at the spheno-ethmoidal and spheno-occipital synchondroses, which ossify at seven years old and between 13-17 years old respectively (Melsen 1972). The viscerocranium is composed of the: frontal, sphenoid, ethmoid, lacrimal, zygomatic arches, maxilla, vomer, palatine and mandible bones. The orbits, nasal cavity, oral cavity as well as the frontal, ethmoidal and maxillary sinuses are all located in the viscerocranium (Moore and Dalley 2006).

Bjork (1963) used implants at sites known to be stable throughout development and radiographs to investigate facial growth. This longitudinal study on individuals aged 4 to 24 years old found that growth is rotational and highly individual. Björk and

Skieller (1983) suggested that the increase in height of the maxillary complex occurred through sutural growth and selective resorption/deposition, whilst the width is increased via the palatal suture. They also suggested that growth occurs in a downward and forwards manner. However, this depends on the relationship between growth rates anteriorly and posteriorly. The direction of the rotation subsequently determines the skeletal pattern of an individual, which was summarised by Solow and Houston (1988).

2.3 GROWTH THEORIES

Historically, craniofacial growth has been described in relation to six theories. The remodelling theory suggests that growth is due to patterned resorption and deposition of bone, with no influence by the sutures or cartilage (Brash 1924). This is likely to be partially true. However, the importance of primary cartilage has been proven in research published after this by Petrovik et al. (1975). The suture theory suggests that sutures and cartilage generate the force for growth (Weinmann and Sicher 1947), whilst the cartilaginous theory is based on the importance of cartilage driving growth in a particular direction, not sutures (Scott 1953; Scott 1954,1956). The concept of sutures being a passive structure is more widely accepted in current practice as transplanting a suture produced no further growth (Ryöppy 1965). The concept of cartilage driving growth is more complex, but there is evidence for nasal septum cartilage (Copray 1986) and the sphenooccipital synchodrosis (Copray and Duterloo 1986) producing growth when transplanted. However, condylar cartilage does not appear to have an influence on growth (Rönning and Koski 1969). Sutures and cartilage appear to be under genetic control (Rönning and Koski 1969).

The functional matrix theory places importance on the interaction between periosteal matrices (muscles and tendons) and capsular matrices (brain and eyes) on growth (Moss and Salentijn 1969). They suggested that the pressure from the soft tissues influence growth rather than growth being genetically influenced. This is likely to be partially true, particularly as an underdeveloped orbit is produced when an eye is congenitally missing. However, the presence of any orbit must suggest that the eye itself is not the only factor involved in the orbit development.

The part-counterpart principal described by Enlow (1990), mapped the craniofacial region into areas anterior and posterior to the posterior maxillary plane. The interaction between these areas vertically, produces balanced growth, allowing all aspects of the head to function appropriately. The complex interaction between the mandible and maxilla supports this, although other elements are likely to be involved.

Finally, the servosystem theory (Petrovik et al. 1975; Petrovik et al. 1981) proposes that growth of the midface occurs due to primary cartilages of the cranial base and nasal septum, which are genetically controlled. The mandible subsequently adapts to its changing position by growth at the condyle and adaption of muscles and soft tissue. Lavergne and Petrovic (1983) subsequently attempted to map out all possible influences on facial growth. Given more recent research into the influence of genetic and environmental factors on facial shape, as well as the potential for epigenetic influences, it is unlikely that any of these theories describe the influences of facial development in sufficient detail.

2.4 THE INFLUENCE OF GENETICS AND ENVIRONMENTAL FACTORS ON FACIAL SHAPE

Genetically, homeobox genes appear to be important and are observed in animals (Quinonez and Innis 2014). Previously, it was suggested that intramembranous ossification was controlled by the transcription factors *MSX-1* and *MSX-2* (Ferguson 2000), whilst Indian Hedgehog (*IHH*) (St-Jacques et al. 1999) and Fibroblast Growth Factor Receptor 3 (*FGFR3*) are important for regulating endochondral ossification (Colvin et al. 1996). More recently, as summarised by Claes and Shriver (2016), genome-wide association studies (GWAS) have been conducted which have found associations between facial shape and single nucleotide polymorphisms (SNPs) (Cole et al. 2016; Shaffer et al. 2016). A summary of the genes known to influence facial features in populations without craniofacial dysmorphia up to 2018 was suggested by Richmond et al. (2018) (Figure. 1). However, more recent GWAS studies have suggested 203 genomic regions (White et al. 2021) and 472 genomic loci (Naqvi et al. 2020) are associated with “normal” facial variation. The genetic control surrounding “normal” facial variation is therefore likely far more complex than was previously thought.

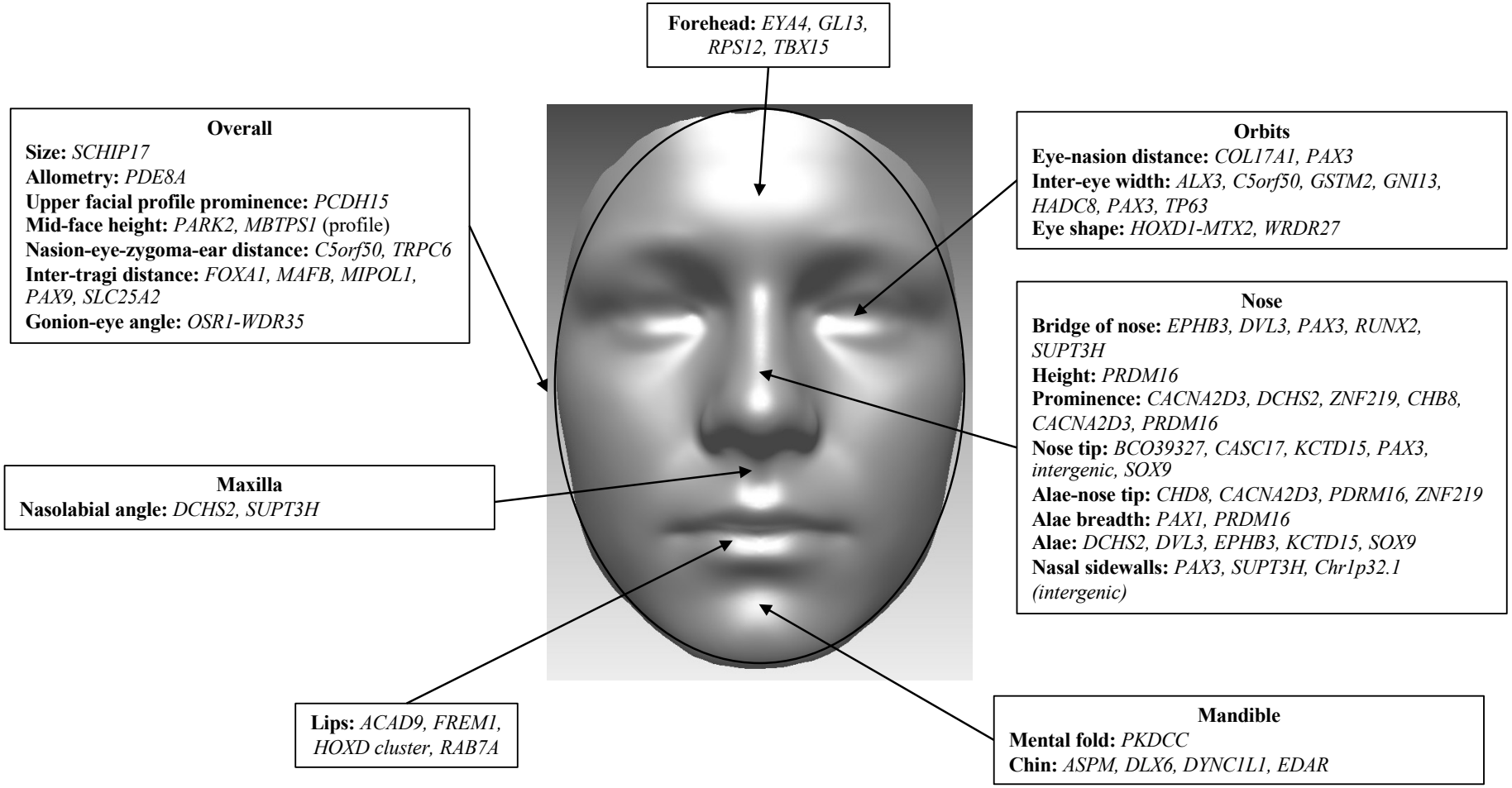


Figure 1: A summary of the influence of genes on the face (reproduced from (Richmond et al. 2018) with permission from the author who is the copyright holder)

Previously, twin studies and family models had been used to investigate the role of genetics in facial development, with most of these using 2D images (Manfredi et al. 1997; Savoye et al. 1998; Carels et al. 2001; Peng et al. 2005; Baydaş et al. 2007; Amini and Borzabadi-Farahani 2009; Alkhudhairi and Alkofide 2010). Other studies have used 3D facial scanning to investigate twins (Weinberg et al. 2013; Djordjevic et al. 2016) with Djordjevic et al. (2016) suggesting that 70% of phenotypic facial differences can explained by genetics. It therefore follows that 30% of facial differences may be explained by environmental factors. However, the twins-model may not be as reliable a method as previously thought for assessing genetic influences. Jonsson et al. (2021) suggest that early developmental mutations can occur, with 15% of monozygotic twins presenting with mutations that are different to those of their twin. Monozygotic twins are therefore unlikely to present with identical genetic profiles, thus reducing the ability of the twins-model to separate environmental and genetic effects.

A further paper investigated the heritability of facial features from fathers to offspring. They used 7160 quasi-facial landmarks and divided the face into regions to assess whether these regions showed low, moderate or high levels of heritability. They found that the upper face including the forehead, nose and cheeks showed higher levels of heritability compared to the lower face, particularly for boys (Hoskens et al. 2018).

Environmental factors such as trauma, surgery, smoking, mastication, bruxism, swallowing, hormone levels, nutritional status, disease states and breathing could all influence facial growth. However, due to the multifactorial nature of a patient's environment, this is perhaps the most complex area to investigate reliably. It is also incredibly challenging to disentangle the influence of environmental factors from genetic factors, particularly with the growing body of evidence around shared genetics between different factors; some of which may previously have been attributed to environmental factors in isolation. For example, it may at first glance be assumed that an increased BMI is due to nutritional status in isolation, which in turn may increase the size of the face. However, BMI is far more complex than this. The Genetic Investigation of Anthropometric Traits (GIANT) consortium aims to investigate genetic loci which influence body shape and size, including height and weight. This

effort has found an association between hundreds of loci and human size and shape (Giant Consortium 2019).

Furthermore, environmental factors may have an influence on genetic factors through epigenetics. Epigenetics centres around the differences between organisms that are not explained by differences coded in DNA (Simmons 2008). These variances are thought to develop through DNA methylation (a methyl group associated with an environmental factor binds to DNA, influencing the expression of a gene) and histone modification, which can lead to silencing of genes or changes to gene expression (Egger et al. 2004).

2.5 THE INFLUENCE OF SPECIFIC VARIABLES ON FACIAL SHAPE

A structured approach was used when practical for conducting this section of the literature review. However, given the range of analytical techniques used in the literature, the heterogeneity of previous studies, the number of variables assessed in this thesis and the limited amount of information available for many of the variables, a narrative presentation was deemed most meaningful.

2.5.1 Geographical area

The facial data used in this thesis is for subjects recruited from populations on the European continent: Croatia, England, Wales and Finland. Previous studies have also investigated the differences between European facial shapes and found subtle differences in most of the main facial features. In general, the differences appear to be very small, in the region of 0.5mm (Table 1).

Table 1: Studies investigating the influence of ethnicity on the facial shape of European populations

Study	Geographical area	Analysis	Differences
Bozic et al. (2009)	Welsh Slovenian	Average faces	Females: Differences in eyes, zygomatic region, mandible except chin prominence Males: Eyes, pronasale, chin prominence
Hopman et al. (2014)	Dutch British	Average faces Heat/colour deviation maps	Females: Dutch have increased facial length, shorter nose, wider nose and wider eyes. Males: Shorter nose and wider eyes. Differences approx. 1mm.
Farnell et al. (2017)	Croatian English Welsh Finnish	Multilevel PCA (2-level)	Populations were differentiable from each other with similar vectors between the male and female means for each population
Kau et al. (2010)	Hungarian Welsh American Slovenian Egyptian	Average faces	Mean differences in European populations were in the region of 0.5mm. Differences in eyes, nose, lips and forehead to varying degrees depending on the population assessed.

2.5.2 Sex

For the purposes of this thesis, sex refers to the sex at birth. The influence of sex on facial shape has been well documented and is in part due to the presence of androgens. Testosterone stimulates the secretion of growth hormone (GH), thus influencing growth (Silva et al. 1992). Oestrogen influences growth in both males and females by stimulating epiphyseal fusion (Weise et al. 2001), thus it is important for normal growth and maturation (Lee and Witchel 1997). A recent study investigated the influence of genetic variants associated with testosterone level on the facial shape of three population cohorts. They found that the testosterone related variants were associated with mandibular shape and facial height to width ratios, which are in turn are related to sexual dimorphism in facial shape (Roosenboom et al. 2018). The sex hormones are rarely found free in the blood; instead, they are bound to glycoproteins such as sex hormone-binding globulin (SHBG) (Selby 1990). It is suggested that GH (De Moor et al. 1972), insulin (Plymate et al. 1988), dietary lipids (Reed et al. 1987) and BMI (Glass et al. 1977) reduce SHBG concentration and may therefore act as confounding factors.

There are a growing number of studies which have assessed the influence of sex on the facial shape of European populations. The findings of these studies are summarised in Table 2. These studies all use different imaging techniques and analyses, as well as assessing different populations and age groups. However, there is general agreement with regards to sexual dimorphism in the craniofacial region. Areas of contention appear to be with regards to the prominence of the mandible. These differences may be due to studies investigating populations that show different levels of dimorphism or different age groups. Indeed, the influence of sex becomes more apparent after puberty (Abbas et al. 2018). A recent study suggests that the facial features influenced by sexual dimorphism change with age. In young children, eye fissure inclination was different between females and males, whilst mandibular position only differed after puberty (Kesterke et al. 2016).

Table 2: The influence of sex on the facial shape of European populations

Facial Difference	Factor	Studies
Overall size	- Females: Smaller with reduced facial height, wider, flatter - Males: Larger with increased facial height	Bugaighis et al. (2013) Ferrario et al. (1996) Ferrario et al. (1999a) Ferrario et al. (1999b) Ferrario et al. (2003) Hennessy et al. (2006) Nute and Moss (2000) Velemínská et al. (2012)
Brow-ridge	- Females: Less prominent - Males: More prominent	Bozic et al. (2009) Ferrario et al. (2003) Gor et al. (2010) Kau et al. (2006) Kau et al. (2010) Koudelová et al. (2015) Mydlová et al. (2015) Nute and Moss (2000) O'Toole et al. (1997) Velemínská et al. (2012)
Orbits	- Females: More prominent, more lateral - Males: Less prominent, larger	Bozic et al. (2009) Bugaighis et al. (2013) Ferrario et al. (2001) Gor et al. (2010) Hennessy et al. (2002)
Cheeks	- Female: More prominent - Males: Less prominent	Bozic et al. (2009) Bugaighis et al. (2013) Gor et al. (2010) Mydlová et al. (2015)
Nose	- Females: Smaller and less prominent - Males: Larger and more prominent	Bozic et al. (2009) Ferrario et al. (1997) Gor et al. (2010) Hennessy et al. (2002) Kau et al. (2006) Koudelová et al. (2015) Mydlová et al. (2015) Nute and Moss (2000) Velemínská et al. (2012)
Lips	- Females: Fuller (especially upper lip), smaller overall size - Males: Less full, larger overall size, thinner upper lip	Ferrario et al. (2000) Ferrario et al. (2009) Hennessy et al. (2002) Petleshkova et al. (2013)
Mandible	- Females: Less prominent - Males: More prominent, wider, profile more inferior and anterior (Some disagreement, perhaps depending on the population or the age of subjects)	Bozic et al. (2009) Ferrario et al. (2003) Gor et al. (2010) Kau et al. (2006) Koudelová et al. (2015) Mydlová et al. (2015) Nute and Moss (2000) O'Toole et al. (1997) Velemínská et al. (2012)

2.5.3 Age and pubertal stage

The longitudinal influence of age is not assessed in this thesis as the subjects studied were between 14-16 years old. (The Croatian population assessed were older, with these subjects used to highlight strengths of the analytical techniques, rather than to explicitly model the influence of age). It should however be noted that 14-16 years old is a crucial time with regards to pubertal development, particularly in boys (Richmond et al. 2020). Pubertal stage was therefore assessed in this thesis. Growth at puberty is determined by a complex interaction of genetic and environmental factors, which are explained by Soliman et al. (2014). The time of onset, duration, amount of growth and time of termination of pubertal growth all vary between individuals. Growth during puberty is controlled by positive and negative feedback loops linking neuro-kinin B, kiss 1 neuron arcuate, gonadotropin-releasing hormone secretion from the hypothalamus, pituitary release of luteinising hormone, follicle stimulating hormone, and sex steroid secretion from the gonads. Levels of insulin, Insulin-like growth factor 1 (IGF-1) and their interaction with GH are also implicated in the process as well as thyroid hormone. Soliman et al. (2014) also suggest that endocrinology pathways are further complicated by the influence of fat mass and energy balance on leptin, which in turn influences the hypothalamus and pituitary gland. It appears that the association between obesity and pubertal growth may be under genetic control (Cousminer et al. 2013).

2.5.4 Height and growth hormone

The height of an individual is driven by genetic and endocrinological factors (Giant Consortium 2019). GH has an important influence on somatic growth (Isaksson et al. 1987), thus it follows that they are likely to influence growth of the craniofacial region. GH is involved in a complex interaction with IGF-1 as part of the GH-IGF axis (Martinelli et al. 2008). Children with normal BMI and reduced levels of GH have been found to present with a reduced anterior cranial base length, increased facial height and increased mandibular plane angle (Kjellberg et al. 2000). Others have suggested that a smaller facial width and reduced posterior facial height with normal anterior facial heights are also associated with GH deficiency (Pirinen et al. 1994). The maxilla appears to be affected less than the mandible, however both have been suggested to be retrusive and hypoplastic when GH is reduced (Kjellberg et al. 2000).

Conversely, excessive GH is associated with increased facial height, increased facial width, and increased clivus (bone behind sella turcica) length (Pirinen et al. 1994). Children with a short stature, but without GH deficiency, have been found to have a similar growth pattern to those with a GH deficiency (Van Erum et al. 1998; Kjellberg et al. 2000).

2.5.5 BMI

Somatic height is increased in obese children (Vignolo et al. 1988). However, levels of GH are reportedly reduced in obesity and normalised when BMI is reduced (Argente et al. 1997). The finding that obese children are taller in the presence of reduced GH levels has been attributed to increased levels of growth hormone binding protein (GHBP) (Hochberg et al. 1992; Argente et al. 1997). However, to further complicate the picture, obese adolescents have been reported to have somatic heights within normal limits (Vignolo et al. 1988).

Subsequently, studies have investigated the influence of obesity on craniofacial development in adolescents. Obese adolescents have been found to have wider faces (Ferrario et al. 2004), elongated anterior cranial bases (Ohrn et al. 2002), increased mandibular widths and lengths (Ohrn et al. 2002; Ferrario et al. 2004; Sadeghianrizi et al. 2005), and decreased mandibular plane angles (Ohrn et al. 2002; Sadeghianrizi et al. 2005). It has been suggested that in individuals with increased BMI, GH has a greater influence on the mandible compared to the maxilla due to the presence of IGF-1 receptors in the temporomandibular joint (TMJ), and an increase in free IGF-1 (Ohrn et al. 2002). There is some disagreement about whether obese faces are retrognathic (Ferrario et al. 2004) or protrusive (Sadeghianrizi et al. 2005), and whether anterior face height is reduced (Ohrn et al., 2002; Ferrario et al., 2004) or increased (Sadeghianrizi et al. 2005), which may be due to the different ethnicities investigated or analyses used.

Female facial shape appears to be more affected by obesity than males (Ohrn et al. 2002; Ferrario et al. 2004). Adipose tissue has been found to have contrasting effects on males and females, with early onset puberty in females occurring with high adipose tissue levels (Wang 2002; Ong et al. 2009) and early onset puberty in males with

reduced adipose tissue levels (Lee et al. 2010). It is therefore important to investigate sexual dimorphism and pubertal status in analyses.

2.5.6 Metabolic factors

The influence of insulin on growth is complex and is determined by the interaction between insulin and GH. Qiu et al. (2017), suggests that the interaction between GH and insulin is important at three distinct stages: 1) before GH release by inhibiting GH producing cells (somatotropes) in the anterior pituitary gland; 2) during the binding of GH to its receptor; and 3) during the cell signalling cascades after GH has bound to its receptor. They suggest that in non-obese and non-diabetic subjects, insulin and GH are regulated so that GH is released at optimal levels. However, they advise that the mechanism of influence of insulin on GH is different in obese individuals, where obesity lowers insulin sensitivity, leading to high levels of insulin in the blood, causing somatostatin (GH inhibiting hormone) to increase, which subsequently decreases GH secretion. In patients with diabetes, high levels of insulin in the blood desensitises the receptors on the GH releasing cells, which increases the level of somatostatin (GH inhibiting hormone) and results in reduced GH release.

There appears to be only one study that has explored the influence of insulin on facial shape. Djordjevic et al. (2013b) investigated the influence of metabolic factors on facial shape using conventional PCA and multivariate regression. They assessed fasting insulin, glucose, total cholesterol, triglycerides, high density lipids (HDL) and low density lipids (LDL) and adjusted for age, sex, pubertal stage and BMI. They did not scale the faces, thus assessed shape and size. They found that insulin was linked with four principal components (PCs) which they state explain facial height, asymmetry of the nasal tip and columella base, asymmetry of the nasal bridge and depth of upper eyelids. Triglycerides were linked to nose height and LDLs were linked to prominence of nose, prominence of lower lip, asymmetry of nasal tip and asymmetry of columella. Notably, none of the associations were significant after Bonferroni correction. However, this may be too conservative given the large reduction in *p*-value required to gain significance in multivariate analyses. Fasting glucose was not found to significantly related to any of the PCs, however, fasting

glucose levels have previously been linked to GH (Riedel et al. 1995), thus further investigation of glucose may also be of benefit.

Further studies have investigated the influence of diabetes on facial shape. Demayo et al. (2009) used two dimensional (2D) photographs and geometric morphometrics to assess the differences in facial shape between non-diabetics and diabetics in 18 to 60 year olds. They found that diabetic patients had facial asymmetry, facial features closer to the centre of the face and more inferior brow-ridges as well as hooded eyelids. Significant flaws in the methodology include the combination of type I and type II diabetics, no mention of the control of diabetes or compliance with insulin regimes and the use of 2D images. This study does however suggest that further investigation into the influence of disorders of a metabolic nature on facial shape may be worthwhile.

Nunes et al. (2018) analysed 2D images of an elderly population. They used DFA with cross-validation and were able to find significant differences in facial shape with regards to diabetes and hypertension. Males were classified correctly more frequently than females suggesting male facial shape is influenced by diabetes more than female facial shape. In frontal view, 75.2% (with mouth and jaw included) and 72.4% male subjects (with mouth and jaw excluded) were correctly categorised as diabetic or non-diabetic. Females were classified correctly 62.8% (with jaw) and 71.4% (without jaw) of the time. In profile, 81.2% (with jaw) and 66.3% (without jaw) of the male subjects were correctly classified. Females were correctly classified 68.5% (with jaw) and 67% (without jaw) of the time. They used thin-plate spline to determine that diabetic individuals presented with increased facial width in the zygomatic region, shorter noses and slight differences in ear form. However, there are outstanding questions with regards to the methodology. It is not stated what type of diabetes was investigated, however, given the age range of the subjects it is assumed the study relates to type II diabetics or latent autoimmune diabetes in adults (LADA) which has features of type I and type II diabetes (Cousminer et al. 2018). The BMI of the subjects was also not considered. As increased BMI is associated with type II diabetes the increased facial width may in fact be due to increased adipose tissue in the cheeks. Furthermore, there is no mention of how and when diabetes was diagnosed. Those with a life-long diabetes diagnosis may be expected to have greater facial shape

differences to non-diabetics than those who developed diabetes in later life. There is also no mention of any medication being taken by the subjects, which may reduce the impact of the insulin resistance associated with type II diabetes. Nevertheless, there is a suggestion that diabetes may influence facial shape.

2.5.7 Breathing disorders and atopy

The influence of breathing disorders on facial shape is controversial. It has been suggested that those with obstructed airflow through the nasal cavity attempt to increase air intake by mouth breathing. This may change the position of the musculature and soft tissues, and facilitate a downward and backwards growth rotation, with increased anterior lower face height (Hannuksela 1981; Harari et al. 2010). This is supported by the work of Linder-Aronson (1974), who reported that following removal of the adenoids in children, significant dental and skeletal changes were seen, alongside a switch from mouth to nasal breathing. However, cause and effect are difficult to ascertain when investigating breathing difficulties with facial growth, particularly as breathing difficulties can arise through multiple aetiologies.

Atopy is a condition defined by the production of immunoglobulin E antibodies to common allergens (Jarvis and Burney 1998). This chronic condition can include suffering from asthma, fever, eczema, utricaria and food allergies, thus may chronically disrupt nasopharyngeal membranes and encourage mouth breathing (Al Ali et al. 2014a). Data on the diagnosis of atopy is available for the ALSPAC cohort at 7.5 years old (Henderson et al. 2008; Al Ali et al. 2014a) and has been linked to increased total face height and mid-face height (Al Ali et al. 2014a). As part of the conditions associated with atopy, asthma affects breathing in a chronic manner and has also been found to be associated with reduced mid-face height in the ALSPAC population, but only in females (Al Ali et al. 2014b). The subtle differences in the associations between atopy and asthma on facial shape are interesting, and thus are worth further investigation. Finally, it has been shown that an increased number of respiratory infections might be linked to facial asymmetry (Thornhill and Gangestad 2006). Although respiratory infections are not assessed in this thesis, this finding provides further evidence for the potential link between breathing issues and facial shape.

2.5.8 Maternal smoking during pregnancy

The likelihood of maternal smoking disrupting foetal development has been investigated for many decades. Studies from the 1970s and 1980s suggested that maternal smoking affects foetal development by generating a state of hypoxia, with (Longo 1976) reporting carbon monoxide as the main teratogen. Others have suggested that nicotine was the main reason for developmental abnormalities after comparing the influence of standard and nicotine-free smoke in rhesus monkeys (Socol 1982). It has also been found that the foetus is exposed to higher levels of nicotine during pregnancy than the mothers, with nicotine concentrations in the placenta, foetal serum and amniotic fluid measured by Luck et al. (1985). Furthermore, it has been suggested that reduced levels of nitric oxide (a vasodilator) in smokers results in relative vasoconstriction, thus potentially reducing oxygen and nutrient availability to the foetus (Andersen et al. 2004). Therefore, there is the potential for maternal smoking to affect foetal development.

There has been an abundance of studies investigating maternal smoking as a potential risk factor for developmental conditions of the craniofacial region. These conditions include reduced head size, craniosynostosis, cleft lip with or without cleft palate (CL/P) and facial asymmetry. Although subjects with any obvious craniofacial dysmorphology were excluded from this thesis, it is feasible that subjects could present as a very mild case, with sub-clinical features, thus an understanding of the influence of maternal smoking on the phenotype of these conditions could be advantageous in understanding non-syndromic variation.

Head circumference at birth has been investigated primarily due to the potential for it to be a good predictor of any neurological development problems (Lipper et al. 1981). Källén (2000) assessed the influence of maternal smoking on the head circumference of infants in a Swedish population and found that head circumference was reduced with increased levels of maternal smoking, despite correcting for reduced overall size of the infants. The smoking groups used were any smoking, smoking <10 cigarettes per day, and smoking 10+ cigarettes per day, with these groupings determined via interview at 10-12 weeks of pregnancy. A dose-dependent relationship was suggested as the odds ratios were 1.48 and 1.74 for <10 cigarettes per day and 10+ cigarettes per

day respectively. Unfortunately, information is only available on the number of infants with head circumference <32cm or 32+ cm, thus it is difficult to assess clinical significance. Clinical significance is important to assess as other studies have reported a reduction in head circumference with maternal smoking of the order of 0.12 +/- 0.03cm (Wang 1997). The influence of smoking cessation during pregnancy was also not assessed by Källén (2000). Lindley (2000) subsequently investigated maternal smoking cessation in a proportion of the same cohort from Sweden. It was found in this study that if mothers stopped smoking by the 32nd week of pregnancy, the head circumference of infants whose mothers stopped smoking was comparable to those whose mothers did not smoke. The importance of stopping smoking was also found in a robust study of a Danish population, with this attributed to reduced nitric oxide levels in the smokers compared to those of non-smokers and ex-smokers (Andersen 2009).

However, it could be argued that any cranial differences exhibited by the infants of smokers and non-smokers could reduce throughout childhood as the child is exposed to post-natal environmental factors. It is therefore important to consider the long-term influences on children of mothers that smoke during pregnancy. Fried (1997) assessed the influence of maternal smoking on children aged one to two years old and found that head circumference was reduced in the children of smokers. However, this effect was most noticeable when mothers had continued to smoke into the third trimester. Furthermore, Kozieł et al. (2018) investigated the longer-term influence of smoking of mothers and fathers during pregnancy in a Polish population of 7 to 10-year-olds. It was found that head length was affected in the male subjects, and not the females. However, the male cephalic index (the ratio of head width to length) showed little difference between the parental smoking and non-smoking groups by 10-years-old. There is therefore evidence that maternal smoking may be a risk factor for reduced head circumference, particularly if mothers smoke for the duration of their pregnancy and if their child is male, although the long-term influences on head circumference and associated clinical significance are still in question.

Craniosynostosis is a developmental condition defined by the premature fusion of at least one of the cranial sutures (Flaherty et al. 2016). Craniosynostosis can influence facial size and shape depending on the sutures that fuse prematurely. Many studies

have looked at the potential influence of maternal smoking on the development of this condition. From these studies, the heterogeneity in the literature is clear because different smoking levels and pregnancy statuses were examined. However, there is some evidence that maternal smoking may influence craniosynostosis. There may be some evidence of a dose-dependent relationship (Alderman et al. 1994; Carmichael et al. 2008). There is also evidence that males are more affected by craniosynostosis than females (Källén 1999). Furthermore, there appears to be disagreement as to which sutures are more susceptible to premature closing with one study suggesting that the sagittal suture craniosynostosis was strongly associated with maternal smoking (Källén 1999) and a second study reported that the coronal suture had a stronger association with maternal smoking (Alderman et al. 1994). Approaches have included broad categories of non-smoker against smoker (Woods and Raju 2001) or <10 cigarettes per day and 10+ cigarettes per day (Källén 1999), whilst others have distinguished smoking levels into smaller categories of up to 20+ cigarettes per day (Alderman et al. 1994; Wang et al. 1997). This makes studies heterogenous and may explain the different conclusions regarding the influence of maternal smoking on developmental conditions. A recent systematic review conducted by Hackshaw (2011) found that maternal smoking is a risk factor for craniosynostosis. Five studies were included in the analysis with a reasonable amount of homogeneity found between the studies (I^2=51%, p=0.09). The authors therefore used random-effects meta-analysis to calculate that the risk of craniosynostosis is increased with maternal smoking, which gave an odds ratio from meta-analysis of 1.33 (95% CI: 1.03-1.73). However, this finding should be interpreted with caution given the exclusion of non-English articles, the small number of studies, the lack of funnel plots, the variability in clinical presentation of craniosynostosis and the heterogeneity of the included studies.

Facial asymmetry is another aspect of craniofacial shape and size that has been researched with regards to maternal smoking. Craniofacial microsomia is a condition affecting the structures formed by the first and second brachial arches. It results in hypoplasia of one side of the face and the associated ear, which are formed during the first six weeks after conception (Birgfeld and Heike 2012). Many classification systems have been developed, with a widely used system being the orbit, mandible, ear, nerve and soft tissue (OMENS) classification, which allows documentation of the

severity of each feature individually (Gougoutas et al. 2007). The wide range of clinical manifestations can be seen and a lack of family history in some cases potentially indicates an environmental aetiology (Werler et al. 2009).

It was initially suggested that a haematoma in the stapedial artery was the cause of craniofacial microsomia (Poswillo 1973), and this has led researchers to investigate other potential factors that could disrupt the blood flow on the affected side. Werler et al. (2009) investigated the influence of maternal smoking on hemifacial microsomia in Canadian and United States of America (USA) populations and found that 1-10 cigarettes per day increased the risk of hemifacial microsomia by 2.3 times but smoking 10+ cigarettes per day was not linked to the condition. The mothers who smoked and took medications in the first trimester that influence the vasculature (non-steroid anti-inflammatories, pseudoephedrine and phenylpropanolamine) were 4.2 times more likely to have a child with hemifacial microsomia. Moreover, previous research on the twins of the ALSPAC cohort suggested that environmental factors may influence asymmetry in regions of the face in different proportions (Djordjevic et al. 2013a), whilst Pound et al. (2014) suggested that childhood ill-health was not implicated in the facial asymmetry of the cohort. It could therefore be suggested that an investigation of maternal smoking on the ALSPAC cohort may explain some of the variation in the individual faces.

Many studies have assessed the influence of maternal smoking on developmental diseases, grouped into anatomical systems. One such study by Woods and Raju (2001) assessed a Cincinnati population and included conditions of the head, tongue, ear, nose, mouth and throat. There were no significant links between maternal smoking and the craniofacial regions in their study. However, due to sample size restrictions, they were only able to assess non-smokers against smokers and did not differentiate between smoking in the different trimesters of pregnancy. Christianson (1980) also found no significant differences in the head and neck region, apart from issues affecting sight. The mothers in this study were asked about their smoking habits at the beginning of their pregnancy only and they were categorised into the following groups: never smoked; stopped smoking before pregnancy; smoked 1-19 cigarettes per day; and smoked 20+ cigarettes per day. This data formed part of the Child Health and Development Studies of California. However, it could be argued that maternal

smoking frequency and deliverable nicotine concentrations from cigarettes may have changed considerably since this study as the subjects' smoking levels were measured in 1959 to 1966. A further study by Malloy et al. (1989) used a Missouri population born between 1980 and 1983, and again found no significant influence of maternal smoking on developmental conditions, with an adjusted odds ratio of 0.84 (95% CI 0.68-1.02) for combined developmental defects related to the eye, ear, nose and throat, and an adjusted odds ratio of 0.84 (95% CI 0.68-1.05) for CL/P. However, these reported odds ratios are in relation to non-smokers against smokers only, with no further categorisation. Furthermore, a recent systematic review by Hackshaw et al. (2011) found that maternal smoking is a risk factor for developmental conditions of the face. Thirteen studies were included, with the results of the random effects analysis suggest there is 19% increased odds of an infant having a facial defect if their mother smoked during pregnancy. However, the authors included defects that affect sight only rather than facial shape alone. Furthermore, the results should be interpreted with caution given the exclusion of non-English articles, the small number of studies, the lack of funnel plots, the variability in clinical presentation of craniosynostosis and the heterogeneity of the included studies.

Despite the volume of evidence for the influence of maternal smoking on developmental conditions of the craniofacial region, there appears to be only one human study that assesses the influence of maternal smoking on non-syndromic or "normal" variation in facial shape. In their supplementary material, Muggli et al. (2017) suggested that maternal smoking may have a small influence on the facial shape of one year old children (0.4-0.5mm). The main differences appear to be in the nasal bridge (more prominent, 0.5mm), forehead (less prominent, 0.5mm), lower lip (less prominent, 0.5mm) and chin position vertically (more superior, 0.4mm). The only other study found in the literature involves the influence of hypoxia on the facial shape of chick embryos. Smith et al. (2013) used geometric morphometrics and PCA to assess the variation in facial shape. They found that the control chicks who were exposed to normal oxygen levels varied in size only, whilst those exposed to various levels of reduced oxygen showed differences in the shape of the eyes, maxilla, nasal structures, brain and frontonasal process. The hypoxic chicks were also developmentally delayed, with some evidence of a dose-dependent relationship. However, a spectrum of differences in facial shape was evident in the chicks. The

results of these previous studies indicate that research into the influence of maternal smoking on the facial variation of human populations would be valuable and of interest. This could further our knowledge of the development of the face and the relative influence of environmental factors on human development. This information could also be used to encourage a reduction in maternal smoking.

2.5.9 Maternal alcohol consumption during pregnancy

It has been established that high levels of maternal alcohol consumption are associated with Foetal Alcohol Syndrome (FAS) (Jones and Smith 1973). Subjects present with short palpebral fissures, and a smooth, short upper lip (Riley et al. 2011). There is also an associated with intellectual difficulties (Jones and Smith 1973). A small number of studies have investigated the influence of differing levels of maternal alcohol consumption on non-syndromic facial shape. Muggli et al. (2017) assessed the influence of alcohol on one year old children using a PLSR model. They investigated the influence of low, moderate and high alcohol levels alongside maternal alcohol tolerance. Low alcohol levels influenced the shape of the forehead (more superior and less prominent) and vertical position of the nose. Moderate to high alcohol intake influenced the shape of the eyes, midface and chin, particularly in a vertical direction. Maternal alcohol consumption at binge levels influenced chin prominence. The authors stated that although these differences were below clinical significance, they could aid in the diagnosis of intellectual difficulties of previously unknown origin.

A further study by Howe et al. (2019) investigated the long-term influence of maternal alcohol consumption on facial shape in the ALSPAC cohort at 15 years old. They determined that although there were subtle differences in the facial shape of the subjects whose mothers consumed alcohol during pregnancy, those differences fell below the criteria for statistical significance. It could therefore be hypothesised that the influence of maternal alcohol consumption reduces with age as other factors begin to influence facial shape. It could however be possible that more sensitive tests are required to ascertain subtle differences. A summary of the current knowledge of the influence of the variables discussed above on facial shape can be found in Figure 2.

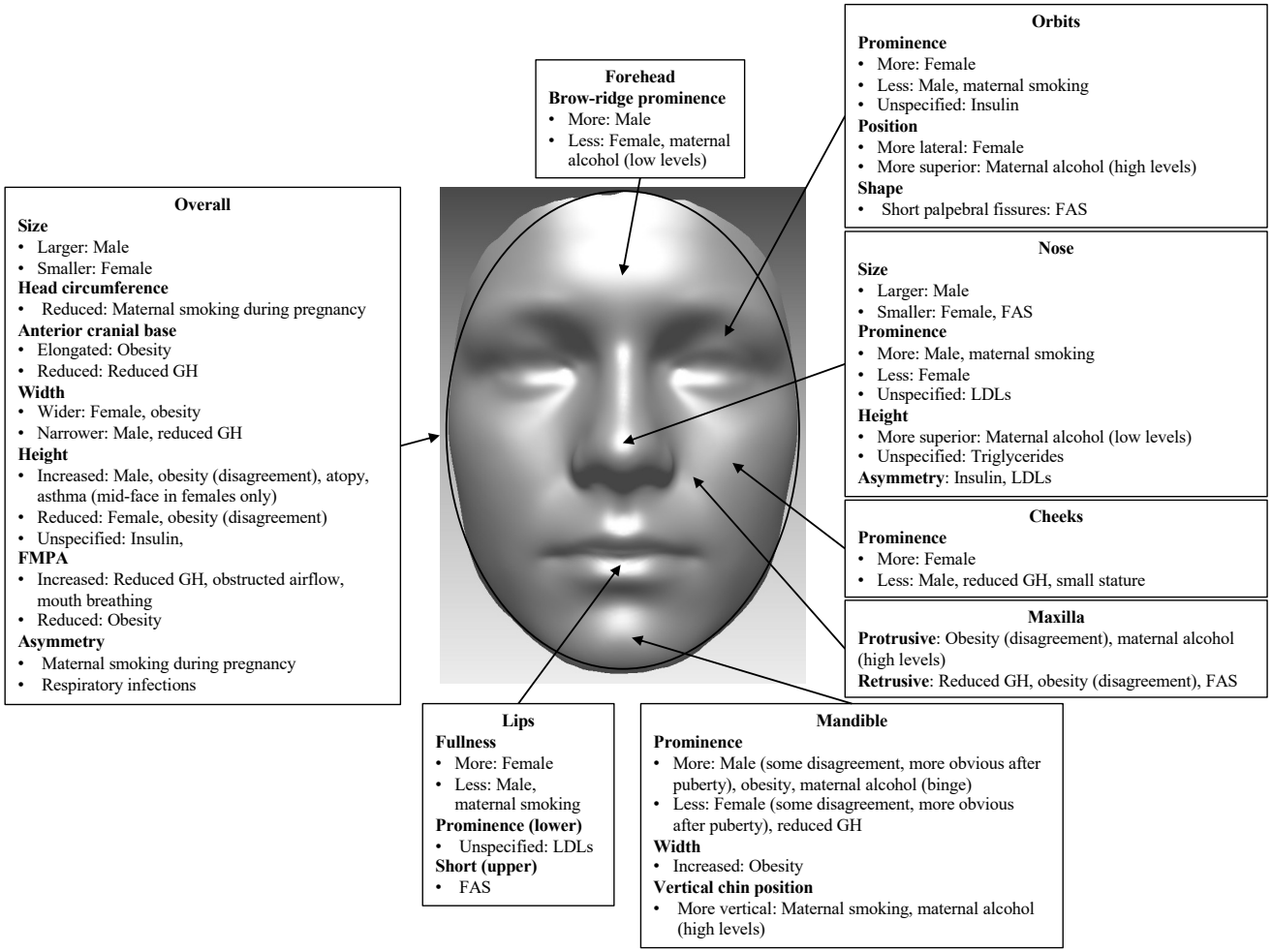


Figure 2: Summary of the findings of previous studies on the influence of the variables investigated in this thesis on facial shape. Growth hormone (GH), low density lipids (LDLs), Foetal Alcohol Syndrome (FAS).

2.6 THE IMPORTANCE OF UNDERSTANDING FACIAL SHAPE

Facial shape is therefore a highly complex area of research with genetic, molecular and environmental factors all likely to have substantial influence on facial development. Encapsulated in facial shape may be an individual's genetics (White et al. 2021) and exposure of their ancestors and themselves to teratogens, pathogens, disease and environmental factors from conception to adulthood (Richmond et al. 2018). Improving our understanding of facial shape is therefore important for many fields including, orthodontics, forensic anthropology, facial recognition and psychology.

Understanding the development of facial shape is essential for orthodontics as it forms an integral part of the orthodontic patient assessment and subsequent treatment options. Currently, an extraoral (facial) examination includes assessment of the patient's antero-posterior profile, with individuals classified into three broad categories. A Class I profile represents well balanced upper and lower jaws, with the mandible placed 2-4mm behind the maxilla. A class II profile exists when the mandible is more than 4mm behind the maxilla. A class III profile represents an individual with the mandible placed less than 2mm behind the maxilla (Cobourne and DiBiase 2016). This may be due to the maxilla being too small, the mandible being too big or a combination of the two.

Transverse discrepancies in the form of facial asymmetry and vertical proportions of the face are also assessed (Cobourne and DiBiase 2016). The extraoral examination is completed by determining the relationship between the lips and nose (nasolabial angle), lip competence (lips meeting or apart at rest) and tooth show on smiling and at rest (Cobourne and DiBiase 2016). This is supplemented by an intraoral examination (assessment of the patient's teeth and bite), with the findings used to determine treatment options (Cobourne and DiBiase 2016). In reality, facial shape is far more complex than this assessment suggests.

Improving understanding of facial shape development could help with one of the most complex issues with regards to determining an orthodontic treatment plan: prediction of how the facial shape will change over time. In a patient with a class II profile, there

is the opportunity to harness the patient's growth potential at the time of the pubertal growth spurt and use an orthodontic appliance that may encourage, to some extent, the mandible to grow forwards (myofunctional therapy). However, there has been controversy around whether this type of treatment causes extra growth, concentrates the patient's growth potential into a smaller period of time, or whether the changes seen are primarily due to tipping of teeth (O'Brien et al. 2003), which is referred to as camouflaging the bite. A more thorough understanding of facial shape and its development could help to determine the true effects of this type of treatment in individual patients and perhaps establish which patient groups are more likely to respond favourably. Of particular importance will be ensuring patients are classified by genotype-phenotype, rather than by an arbitrary overjet threshold.

In patients with growth potential remaining and a class III profile, it can be challenging determine whether braces alone can be used to tip teeth to bring the top teeth in front of the bottom teeth (class I) or whether the mandible will end up too far in front of the maxilla for this to be possible at the end of growth. If camouflage treatment is carried out too early, the patient may grow unfavourably, and present with a suboptimal bite after growth has finished, despite treatment being carried out. This is clearly an undesirable outcome for both patients and clinicians as could result in repeat treatment being required. This is even more objectionable if surgery is carried out too early and the patient requires a repeat operation. Researchers have therefore attempted to find a method of predicting facial growth (for example, (Schulhof et al. 1977; Abu Alhaija and Richardson 2003; Jiménez-Silva et al. 2021). However, there is not currently a method that is considered reliable enough to be certain that unfavourable growth will not occur. This often means that definitive treatment is delayed until growth is finished which requires patients to accept their bite until their late teens or early twenties. Future efforts should focus on matching phenotypes or building more sophisticated statistical models which requires further understanding of the factors involved in facial development. This may aid researchers and clinicians in finding a method of predicting growth, thus determining whether treatment could be carried out earlier in some individuals.

Currently, orthodontic patients are treated to broad population norms. At present, orthodontic treatment is generally centred around active intervention, rather than

taking a preventative approach with regards to development of malocclusion (suboptimal bite). Even interceptive treatment, which is carried out in childhood with the aim of alleviating the potential for orthodontic or dental health problems developing in future (Artese 2019), involves invasive treatment in the form of dental extractions and/or orthodontic appliances. Ultimately, as understanding of facial shape development improves, it may be possible to move away from active intervention in the form of orthodontic appliances, dental extractions and jaw surgery. Instead, it may be possible to influence the development of the face and dentition by adapting genetic or environmental factors and predicting which patients will react most favourably. This concept is used in medicine in the form of personalised medicine (Mathur and Sutton 2017). Aside from a summary of potential areas of interest (Reddy et al. 2019), there does not appear to be any published personalised orthodontic treatments. This is likely because the level of understanding of facial shape in the published literature is currently insufficient to be able to develop a personalised approach to orthodontic care, determine with certainty with patients will respond best to myofunctional therapy or predict which patients will grow unfavourably.

Understanding facial shape is also important for the field of forensic anthropology. Forensic anthropology concerns itself with the identification of individuals (Scheuer 2002). In living individuals, this includes facial recognition. In recent years, more emphasis is placed on facial recognition as technology has improved. In deceased individuals, identifying a body or skeleton can be challenging (Scheuer and Black 2007a). Broad information such as sex, ethnicity, age and stature are all important to help build a picture of the individual (Scheuer 2002). Research has been conducted to find features that could help categorise an individual with the skull seen as the easiest method to determine sex after the pelvis (Scheuer and Black 2007b). Improving knowledge, technologies and analytical techniques could therefore help to further this field.

Given the expressive nature of the face and because the face is visible in most cultures in everyday life, psychologists have also explored facial shape. Malocclusions, and therefore their associated facial shapes, have been associated with psychosocial issues such as being unpopular (Çokakoğlu et al. 2016). Certain facial traits in males have

been associated with behaviours such as aggression (Haselhuhn et al. 2015). Facial shape has also been associated with immunity levels in males due to facial adiposity which may subsequently influence mate preferences (Rantala et al. 2013). However, more recent studies have advised caution over associated facial features with perceived genetic "quality" (Van Dongen et al. 2020).

Furthermore, facial features have been found to be associated with bipolar disorder (Hennessy et al. 2010), schizophrenia (Hennessy et al. 2007) and autism (Tan et al. 2017). Diagnosis of psychiatric and neurological disorders can be challenging. Analysis of the face could therefore screen for and help with the diagnosis of neurological conditions. It may also be possible to use facial shape to screen for other medical conditions and syndromes (Ferry et al. 2014), thus improving analytical techniques may be able to detect sub-clinical features that could indicate potential carriers of genetic conditions. More far reaching, could be the possibility of screening for conditions such as diabetes in a non-invasive manner.

2.7 METHODS OF ANALYSING THE CRANIOFACIAL REGION

Given the number of potential advantages for understanding facial shape, researchers have investigated the craniofacial region extensively using various techniques to assess both hard and soft tissue. These techniques have progressed from measuring distances on the faces of living subjects and on the dry skulls of deceased individuals, to assessing 3D facial images using sophisticated computer algorithms.

Anthropologists have historically assessed individuals in real-time, using landmarks as described by Farkas (1981), as reference points. Subjects are analysed with the Frankfurt plane horizontal. Marks are placed directly onto specific points on the face to make visualisation and consistency as easy as possible (Farkas and Deutsch 1996). To study differences between individuals and populations, the distances between points of interest are often measured. Forty-seven points and 132 measurements have been described Farkas (1994). It is time consuming if a large number of points are used and in living individuals, subject cooperation may be an issue, particularly when measuring anxious or younger individuals (Farkas and Deutsch 1996).

In deceased individuals, the skull may be the only medium available for analysis. The skull is analysed in a similar manner as above, with the distances between points measured as accurately as possible (Scheuer and Black 2007b). This has advantages over working with living individuals, as cooperation is not problematic. However, it is still time-consuming, relies on investigator skill and, most importantly, does not provide direct information on soft tissue profile. Important information is therefore lost. As images and image processing software are not required, the errors involved in image processing are avoided. However, the investigator must be able to reproduce the points accurately, which takes considerable skill. Also, the investigator must be available for a substantial amount of time to measure the skulls, making repeat measurements or checking measurements difficult.

Studying both hard and soft tissue structures of the craniofacial region of living individuals has historically involved lateral cephalometric analysis. These radiographs are taken in a standardised manner, with a collimated x-ray source placed two meters away from the film in the patient's midsagittal plane. With the head horizontal to the Frankfort plane, a cephalostat is used to ensure a reproducible position and an aluminium filter wedge is used to allow soft tissues to be visualised on the image with the hard tissues (Whaites 2002a). A number of standard points are used routinely in orthodontics, with numerous analytical methods available depending on the aspect of the craniofacial region being investigated. The standardisation involved in taking lateral cephalometric images is beneficial as it allows for comparison of different patients and during development of the same patient. There are issues with magnification in the order of 10%. However, the scale included in the image minimises the influence of these issues (Cobourne and DiBiase 2016). As with the measurement of points on skulls and on living individuals, various error rates are found at each of these points (Baumrind and Frantz 1971). A further disadvantage of the use of lateral cephalometric analysis is the exposure of the craniofacial region to radiation. The dose of a lateral cephalogram is 5-10μSv (Hoogeveen et al. 2015) with risks in the craniofacial region particularly important due to the presence of the eyes and brain. Stochastic (random) effects are those effects that may happen when a patient is exposed to radiation and include cancer or genetic effects (Whaites 2002b). Deterministic effects or tissue reactions are effects that will definitely happen if a

patient is exposed to a high radiation dose, and include erythema and osteoradionecrosis (Whaites 2002b). It is therefore not ethical to take these radiographs without clinical justification, and thus it is no longer possible to investigate growth and facial shape in individuals who do not require orthodontic or orthognathic treatment. Also, given the 2D nature of this imaging technique, the information obtained is limited.

Computed tomography (CT) scans can be used to produce 3D radiographic images of individuals, which is beneficial in gaining more robust information on shapes and growth of the craniofacial region. However, the radiation dose of a CT to the maxilla or mandible is 100-3000μSv (Commission 2004), thus the radiation risks discussed for lateral cephalograms are greater. This makes it ethically impossible to justify taking regular CT scans to analyse growth in a healthy individual. The development of cone beam computed tomography (CBCT) has allowed for a reduction in radiation dose to the patient. The dose for a typical CBCT is 10-1100μSv depending on the size of the area investigated (Pauwels et al. 2012). However, the use of radiation still harbours risk. Using no radiation would therefore be preferable.

Non-invasive techniques include ultrasound, Magnetic Resonance Imaging (MRI), 2D photography and 3D scanning. Ultrasound is a non-invasive technique, which emits sound waves of high frequency in thin slices. These are reflected back, through the transducer, and are processed into 3D images by computer software (Papadopoulos et al. 2002). It has been utilised to investigate soft tissue thickness in populations (El-Mehallawi and Soliman 2001). However, distortions and artefacts can make the images inaccurate (Papadopoulos et al. 2002).

MRI uses a magnetic tomograph, which utilises magnetic fields to polarise hydrogen atoms. When these atoms depolarise, they release radio wave-like radiation, which is subsequently received and can be processed to produce 3D images (Papadopoulos et al. 2002). The benefit of using MRI to analyse the craniofacial region is that no radiation is used. It also allows for imaging of both hard and soft tissues, thus can give useful information on the relationship between the two. Furthermore, it provides 3D information on subjects, which allows for more detailed analysis of facial shape

and growth. It is for these reasons that researchers have used MRI to investigate soft tissue thickness in populations (Chen et al. 2011). However, the process of taking an MRI can be difficult for patients, particularly children or those who are claustrophobic. It cannot be used in those with metal incorporated into their body as these produce artefacts and the magnetic field may displace the metal, causing tissue damage (Papadopoulos et al. 2002). Furthermore, it is time-consuming and a costly medium as it requires expensive equipment along with specialist training to take the images (Greene et al. 2016).

Photography provides a simple, cost-effective, and safe method of imaging the craniofacial region. It can be used regularly to allow for comparison of subjects in a longitudinal manner, as well as comparing different individuals. In the field of facial recognition, Chellappa (2005) advise that lighting must be controlled to allow for high quality images to be taken. It is also important that subjects have a reproducible, neutral expression in the photographs to allow for reliable comparisons, unless for analysis of dynamic processes such as smiling (Chellappa 2005). However, ageing and changes to appearance such as make up, may make it more difficult to plot points accurately (Chellappa 2005). As this method only allows for soft tissue analysis, bony growth can only be estimated. Furthermore, although photographs can be taken in the sagittal and coronal planes, it has been suggested that parallax from photographs at different angles can cause distortions (Farkas et al. 1980). Information is therefore lost as the two-dimensional nature limits the amount of information available. However, work has attempted to limit this (Rabey 1971).

A method of gaining three-dimensional information of facial shape involves 3D facial images. The benefits include the efficiency of image acquisition, 3D information and non-invasive nature make this a popular choice going forwards for analysing the craniofacial region (Papadopoulos et al. 2002). Stereophotography can involve the use of two or more cameras to take a picture of the subject's face at different angles, these are then stitched together using specialised software. Recent advances have simplified this process further using one camera, taking pictures at different angles (Camison et al. 2018).

Alternative 3D surface technology involves laser surface scanning to project a beam of laser light across the surface of the face. The distortions in the laser beam caused by the facial features are recorded by computer software and used to create a 3D image (Mah and Hatcher 2003). This has been found to be more accurate than stereophotography (0.3-0.5mm compared to 0.6-1mm) (Kau et al. 2005; Zhurov et al. 2010), while more recent studies have suggested the techniques are comparable (Camison et al. 2018). However, image capture takes longer than stereophotography. For both stereophotography and 3D laser scanning, steps can be seen in the images where the computer software has had difficulty processing the image, whilst shadows and patient movement can also affect the reliability of the image. As with 2D photography, it is important that external factors such as lighting and facial expression are controlled to allow for accurate images to be taken. Issues with facial hair obscuring specific landmarks can also reduce the information available from the images (Farkas and Deutsch 1996). The use of grey-scale images, as opposed to colour images can reduce bias and difficulties plotting points due to factors such as skin texture and make up. As the technology is relatively new, the equipment and associated software have always been expensive (Ward 1994). Although advances in the technology are continually being made to reduce these issues. 3D laser scanning was used to determine the facial shape of the subjects in this thesis.

The advantages and disadvantages of the different imaging techniques are summarised in Table 3.

Table 3: Methods of analysing the craniofacial region

Method	Description	Advantages	Disadvantages
Direct anthropology	• Visual analysis • Frankfort Plane horizontal • Marks placed directly onto face to increase consistency (Farkas and Deutsch 1996). Distances between points frequently measured	• Living individuals therefore modern data • No image processing	• Difficult and time consuming to reproduce points accurately, particularly if subject cooperation challenging (Farkas and Deutsch 1996) • Subject must be able available for considerable time
Dry skulls	• Points placed or visualised on dry skull • Distances between points measured as accurate as possible (Scheuer and Black 2007b)	• No cooperation required • No time constraints with regards to subject availability	• Time consuming • Relies on operator skill • No direct information on soft tissues • Data could be historical if from historical population • Considerable bias possible if information on subject is incomplete (and may be difficult to obtain further information)
Lateral cephalometry	• 2D standardised radiograph in profile • Landmarks placed either manually or automatically and used to determine distances and angles	• Standardised • Relatively low radiation dose	• Ethically challenging to justify radiation dose in a healthy individual without clinical need • Magnification factor of 10% (although scale limits the impact of this) • Error rates associated with landmarks (Baumrind and Frantz 1971) • Superimposition of structures, with particularly difficulty in plotting those not in the midline
CT	• 3D radiograph • Landmarks placed manually or automatically in order to determine shape, distances and angles	• Detailed 3D information on hard and soft tissues	• Ethically challenging to justify radiation dose in a healthy individual without clinical need

CBCT	• Similar to CT radiograph but using a smaller radiation dose	• Detailed 3D information on hard and soft tissues • Less radiation than CT (10-1100μSv depending on the size of the area investigated (Pauwels et al. 2012)	• Ethically challenging to justify radiation dose in a healthy individual without clinical need
MRI	• Magnetic fields polarise hydrogen atoms • The atoms depolarise and release radio wave-like radiation • Processed to produce 3D images (Papadopoulos et al., 2002)	• No radiation • Detailed 3D information on hard and soft tissues	• Process can be difficult for children or those who are claustrophobic • Cannot be used in those with metal incorporated into their body as these produce artefacts and the magnetic field may displace the metal, causing tissue damage (Papadopoulos et al. 2002) • Time-consuming • Expensive equipment • Specialist training needed to interpret images
Ultrasound	• Sound waves of high frequency emitted in thin slices and reflected back, through a transducer. Processed into 3D images by computer software (Papadopoulos et al. 2002)	• No radiation • Information on soft tissue thickness possible (El-Mehallawi and Soliman 2001)	• Distortions and artefacts can make the images inaccurate (Papadopoulos et al. 2002)
2D photographs	• 2D images taken using a digital camera or using film	• Simple • Cost effective • No radiation • Equipment readily available	• Difficult to ascertain what differences are due to hard tissues and which are due to soft tissues • Information in 2D only
Stereo photography	• Photographs taken using a specialist camera at different angles and stitched together using computer software	• No radiation • Information in 3D • Quick to take images	• Difficult to ascertain what differences are due to hard tissues and which for soft tissues • Requires specialist equipment and software
3D laser scan	• Laser scans across surface with distortions due to the facial features recorded	• No radiation • Information in 3D	• Difficult to ascertain what differences are due to hard tissues and which for soft tissues • Takes longer to obtain an image than stereophotography

2.8 COMPARING 3D FACIAL SCANS

There are two broad strategies for the analysis of surface facial shape. The use of the full facial mesh or the use of landmarks. The full facial mesh can be used to determine differences between the facial shapes of different groups via average faces for each group. This method has been used by many research groups (for example, (Kau et al. 2006; Bozic et al. 2009)), with the differences often visualised using coloured heat/deviation maps. Using average faces allows differences to be visualised between-groups but gives no indication of the differences within the groups without additional standard deviations.

The use of landmarks is the focus of this thesis. This involves the placement of landmarks on the scanned surface to determine the position of facial features. One method of placing these landmarks involves manually plotting using computer software. This allows the investigator to have complete control over the positioning of the points. These landmarks can be used directly to compare faces, or distances and angles between these points can be used to compare the facial shape of subjects. However, as 3D laser scans are computer images, reproducibility of the points relies on good anatomical descriptions, the ability of the investigator and good image resolution. This can therefore limit the number of points that can be readily identified.

Plotting points on 3D facial scans can be difficult and time consuming as they must be placed accurately in each of the three dimensions. Aung et al. (1995) compared direct anthropometry to plotted points on a 3D scan and found that only 49.4% of linear measurements from these points were comparable between the two methods. However, improved technology since this study was published may produce more reliable comparisons. A more recent study by Toma et al. (2009) assessed the inter- and intra-reliability of manually plotting 21 points onto 3D facial scans. They found that approximately 50% of points were plotted with <1mm of error. However, they found that the orbit is difficult to plot due to the complex anatomy of the area, which is confounded by the difficulty in the laser beam accurately recording the detailed anatomy in the area (Papadopoulos et al. 2002). This is potentially problematic in this research as the eye could demonstrate differences between populations.

More recent research has used quasi-landmarks which can provide thousands of landmarks for analysis. The process of placing quasi-landmarks used in this thesis was developed at KU Leuven, with the methodology explained by Claes et al. (2012) and outlined in the Methodology chapter (3.5 Landmarking, p. 52).

This thesis begins by using 21 landmarks as described by Farkas (1994) as a proxy for the main facial features. These landmarks serve as a manageable number to test the ability of the analyses to cope with numerous outcome variables and allowed for comprehensible comparisons between different statistical techniques. However, this significantly limits the amount of clinical information available. Latter analyses use 1000 three dimensional quasi-landmarks and 7160 three-dimensional quasi-landmarks to increase the amount of information available.

2.9 ANALYSIS OF 3D FACIAL LANDMARKS

Analysis of differences in the facial shape of subjects can be descriptive in nature and/or involve inferential tests. Both descriptive analyses and inferential tests are used in this thesis.

2.9.1 Multivariate Analysis of Variance

MAVOVA has been used to test if there are significant differences in shapes (represented here by landmark points) between groups. MANOVA is an extension of analysis of variance (ANOVA) (Tabachnick and Fidell 1996), which allows multiple continuous outcome variables (dependent variables) to be compared with multiple categorical variables (independent variables). ANOVA establishes the within-group and between-group variations. These are used to determine the F-statistic, which subsequently allows a *p*-value to be calculated. If the between-group variation is large and the within-group variation small, there is a significant difference between the groups. If the between-group variation is small and the within-group variation large, there is not a significant difference between the groups (Tabachnick and Fidell 1996).

However, MANOVA has several disadvantages within this field. Firstly, MANOVA requires the number of outcome variables to be less than the number of subjects in

each group to be reliable (Tabachnick and Fidell 1996). This is rarely the case in facial shape research, particularly as the number of landmarks increases. Secondly, the power of MANOVA is reduced in the presence of multicollinearity (variables are strongly correlated) (Tabachnick and Fidell 1996), which is common in facial shape research. It can therefore be beneficial to try to reduce the number of outcome variables that require analysing using dimension reduction (Sha et al. 2011). This can also help reduce the need for increased computer power, storage space and time taken to analyse data.

2.9.2 Conventional PCA

A common dimension reduction technique is conventional PCA, which can be used to reduce the original outcome variables (landmarks) into a more manageable number of principal components (PCs) (Jolliffe 2002). These can subsequently facilitate MANOVA. PCs are related to the original outcome variables but must be interpreted afterwards to assess what they mean in the context of the original data. PCs are linear combinations of the original landmarks (Jolliffe 2002). For example, one PC might explain some of the variation in the subjects' nose shape, width of the mouth and size of the eyes.

PCA begins by calculating the direction of the maximum variation in the data set in as many dimensions as there are original variables and is summarised well by Ringnér (2008). If there are 63 original variables (i.e., 21 three-dimensional landmarks), the direction of maximum variation will be calculated in 63 dimensions. This information is summarised as a covariance matrix (grid of numbers) in this thesis. Covariance is the amount two variables vary together (like the height of a person and the weight of a person in a population). Only two dimensions can be visualised properly on a page, as demonstrated in Figure 3. However, the PCA process occurs in as many dimensions as there are outcome variables.

The lines of maximum variation have two properties. The first is the direction, known as the eigenvector. The second is the amount of variation the line explains, which is known as the eigenvalue. In PCA, the eigenvectors are ordered by their eigenvalue. The eigenvectors with the highest eigenvalue are most important as they explain the

most variance in the dataset. The eigenvector with the smallest eigenvalue explains the least variance. Once they are ordered in descending order, they are renamed as PCs. In a model with 63 PCs, PC1 explains the most variance and PC63 explains the least. It is possible to work out how important each PC is in relation to the original dataset (e.g., PC1 explains 85% of the total variation, PC2 explains 12% of the variation and PC63 explains 0.1% of the variation). The PCs that explain a very small amount of variance may either represent noise in the data or may just be seen as "not that important" in comparison to the other PCs. A decision is then made about which PCs to look at further. The PCs that explain the smallest amount of the variation can be discarded (Jolliffe 2002).

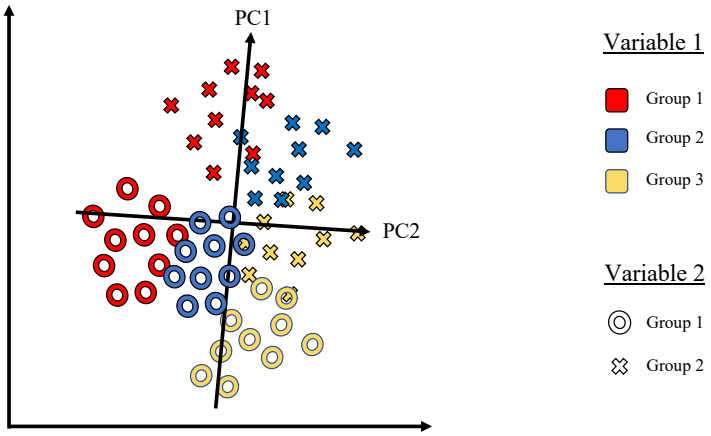


Figure 3: Visualisation of the process involved in PCA. This figure shows that PCA is simply a rotation of axes so that PC1 captures the largest amount of variation in the data (irrespective of any groupings in the data), PC2 captures the second largest amount of variation, PC3 (not shown here) the third largest, and so on.

However, conventional PCA has several disadvantages. The main issue with conventional PCA is that the variables or groupings of subjects are not considered in the dimension reduction process (e.g., sex) (Farnell et al. 2017). There is therefore a risk of losing meaningful information and the interpretation of the results is less clear as the influence of each variable is mixed with the influence of many others in each PC.

2.9.3 Multilevel Principal Component Analysis

Unlike conventional PCA, mPCA takes variables into account when the eigenvalues and eigenvectors are calculated, making it a "guided" dimension reduction technique. mPCA was suggested by Timmerman (2006) and adapted to assess dental radiographs by Farnell et al. (2016), who developed an in-house mPCA code in MATLAB. mPCA can model the maximum variation between groups of subjects which makes it a promising technique in a field where the differences between groups are likely to be subtle. The mathematical techniques are described by Farnell et al. (2016, 2017).

There are three main approaches to mPCA, each applicable for different data structures: nested, non-nested and mixed, which are discussed by Farnell et al. (2020). These methods use slightly different methods for the averaging of the covariance matrices. Utilising a model that fits the natural structure of the data should (a priori) provide a better model than one that does not. A nested model is used when a subject can only belong to one specific group in a particular level. For example, if mPCA was carried out to assess the influence of a child's class and school, a nested model would be most appropriate (level 1: class; level 2: school, level 3: subject). If a child is a member of class one in school A, there is no possibility of them being a member of class one in school B as they cannot be enrolled in two classes simultaneously. The class and the school are linked, and the model is therefore nested. By contrast, a female child could attend either school A or school B, assuming each school is of mixed gender (level 1: school; level 2: sex; level 3: subject). School and sex are not linked in this case; therefore, a non-nested model is most appropriate. If a 2-level model is to be used, this is seen as a special case, and a nested model is most appropriate (level 1: sex; level 2: subject). If repeat measures are carried out on a subject, a mixed approach is most appropriate, for example, if facial scans are taken of a subject over multiple time points (level 1: sex; level 2: subject; level 3: age).

Each variable is assigned a level in the model (e.g., level 1: geographical location and level 2: sex). The within-group variation (all variation not explained by the previous levels) is calculated as a further level (level 3 in this example). In the same manner as conventional PCA, mPCA begins with covariance matrix calculation. In contrast to conventional PCA, a separate covariance matrix is calculated at each level of the

model. At the within-group variation level, a covariance matrix is calculated for each of the groups (i.e., Croatian females, Croatian males, English females, English males, Welsh females, Welsh males, English females, and English males). In this example, eight covariance matrices would be constructed. These are then averaged to give one overall matrix for this level. At the levels focused on between-group variation (i.e., levels 1 and 2), a covariance matrix is constructed using the mean or medians of each of groups with respect to the grand mean. The direction and magnitude of maximum variance is therefore determined between the group means/medians (Farnell et al. 2016; Farnell et al. 2017) (Figure 4).

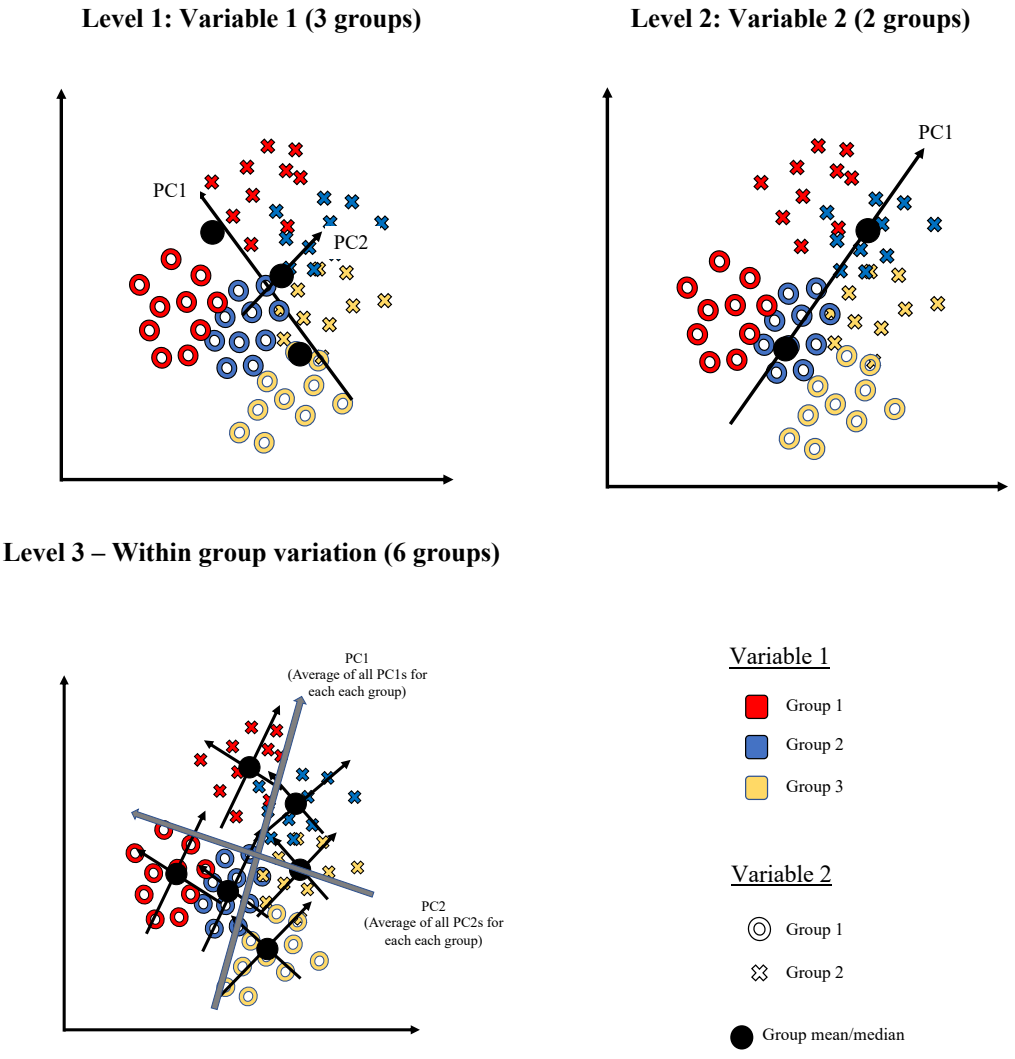


Figure 4: Visualisation of mPCA. A three-level model is described here. Level 1 describes variable 1 (3 groups), level 2 describes variable 2 (2 groups) and level 3 describes the variation within all the groups. At level 1 and 2, the maximum variation between the group means or medians is calculated. At level 3, the maximum variation is calculated for each group. These are averaged to give overall PCs for the within-group variation.

As with conventional statistical techniques, the use of the median can be useful for circumstances where outliers have the potential to impact the results, or the data is not normally distributed. The median of each group was therefore used in this thesis for constructing the between-groups covariance matrices to attempt to control for outlying shapes. A method for dealing with outliers of single landmarks is the use of robust covariance matrix estimation. However, the MATLAB function "robustcov" requires a minimum sample size of greater than 2 x number of dimensions x number of landmarks. In many of the analyses, obtaining a minimum sample size of this magnitude in every group was too restrictive, thus this method was not used.

As per conventional PCA, the PCs with the smallest eigenvalues are discarded. However, when using mPCA, a restriction on the number of PCs is required. For each level, the maximum number of PCs that can be retained is the number of groups minus one, therefore (for example), only one PC can be retained when assessing sex as there are two groups (female and male). The component scores then are calculated by incorporating all the levels simultaneously, resulting in separate component scores for each subject at each level (if there are three levels, one subject will have three sets of component scores).

There are several advantages to using mPCA. Firstly, mPCA can provide clear visualisations in differences between the component scores. This is possible as the groupings are considered during the mPCA process. Further, mPCA can provide an indication of the relative importance of the variables assessed. For example, if level 1 explains 10% of the total variation and level 2 explains 15% of the total variation, there is a suggestion that variable 2 (e.g., geographical location) is more important than variable 1 (e.g., sex) at explaining facial shape variation (Galloway et al. 2020).

Farnell et al. (2017) used mPCA to explore the facial shapes of four European populations: Croatian, English, Welsh and Finnish. They used a two-level model. Level one described ethnicity (Croatian, English, Welsh and Finnish) and sex (female and male). Level two described the within group variation (all the variation that is not explained by level one). They advised that the multilevel nature of the technique provides more flexibility and control than standard models. Clear clustering of the population centroids was seen, meaning that mPCA was able to separate each of the

groups. Of interest was a clear vector between the male and female groups of each ethnicity thus warranting further investigation.

2.9.4 Discriminate Function Analysis

A related technique to mPCA is DFA. This was originally proposed by Fisher (1938) and has subsequently been developed into both linear and quadratic forms. In Linear Discriminant Analysis (LDA) linear equations are generated to separate groups. In Quadratic Discriminate Analysis (QDA) non-linear functions are generated (Figure 5). By calculating the between-group variation and within-group variation, DFA generates equations (or functions) that best separate the group centroids. In LDA, these functions are perpendicular to the lines connecting the group centroids. Scores can be generated by projecting the original landmark data onto this line (Adams, 2005). DFA can subsequently be used to predict information about subjects that were not included in the model initially. This is known as classification and is out-with the scope of this thesis. Like mPCA, for DFA the landmark data must be continuous, but the other variables of interest must be categorical. This is very useful for naturally categorical variables (e.g., sex), but can limit the information possible for naturally continuous variables (e.g., fasting insulin) as categorisation is required.

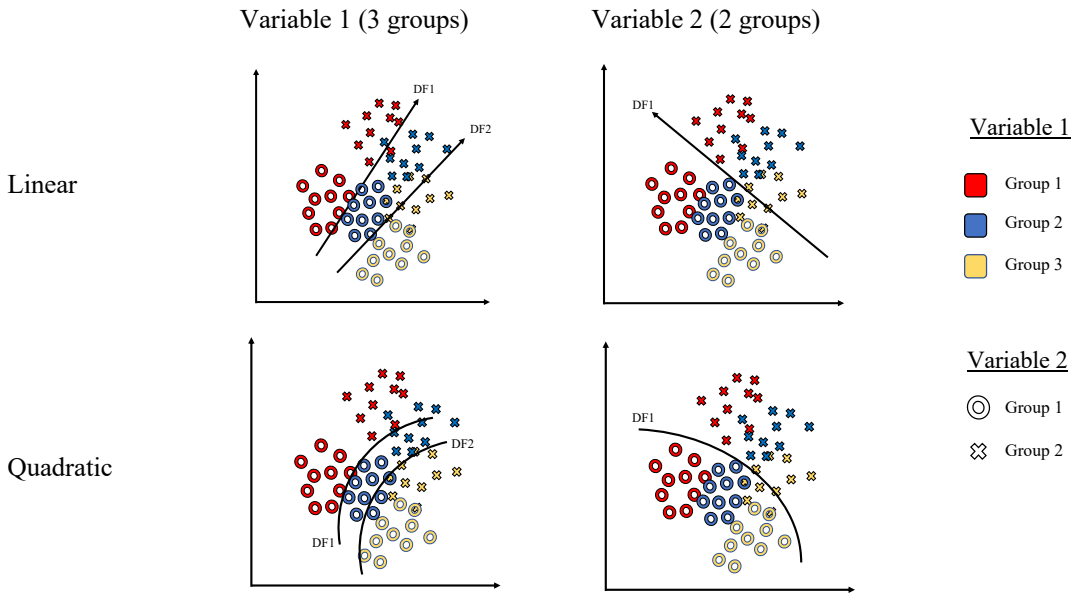


Figure 5: Visualisation of DFA. (a) Linear, (b) Quadratic

2.9.5 Partial Least Squares Regression

In contrast to DFA and mPCA, PLSR is capable of handling continuous variables (e.g., fasting insulin) as well as categorical variables. PLSR was first described by Wold (1966) and is summarised by Geladi and Kowalski (1986) and Wold et al. (2001). It is a supervised dimension reduction technique that takes features from conventional PCA and multiple regression (Abdi 2011). It has been suggested that PLSR is beneficial in circumstances where correlated dependent variables (landmark data) are greater in number than the number of subjects (Shrimpton et al. 2014). It is for this reason that PLSR has been used previously in facial shape research (Shrimpton et al. 2014; Matthews et al. 2016; Muggli et al. 2017). PLSR is also beneficial when the number of independent variables (e.g., sex, height etc) exceed the number of subjects (Abdi 2011, Shrimpton et al. 2014), however, this is not an issue in this thesis.

PLSR calculates the "components" in a different way to PCA. Instead of finding the maximum variation in the landmarks without considering the variables (as in PCA), the components explain as much of the covariance in both the landmarks and the variables as possible. PLSR simultaneously decomposes the predictors (e.g., sex) and response variables (landmarks) with the aim of explaining the maximum amount of covariance between the two (SAGE 2011). Xia (2020) advises PLSR finds the direction in the predictor data that explains the maximum variance in the response data (multidimensional). The PLS components are therefore linear reconfigurations of the initial variables (de Jong 1993) but resolved in a different manner to PCA as both the predictor and response variables are used simultaneously.

As is the norm with the MATLAB function "plsregress", the SIMPLS algorithm is used here to find the weighting matrix required to reflect the covariance between the predictor and response variables (Farnell et al. 2021). SIMPLS was first described by de Jong (1993). In a similar manner to PCA, the first component will explain the highest covariance between the predictors and response variables (Noback et al. 2011). Regression is then performed on the components to predict the variables (Abdi 2011). PLSR is used here as an alternative technique to mPCA due to its ability to handle continuous data and to provide a complementary analysis to further investigate the influence of variables on facial shape.

The list of techniques explored in this thesis is far from exhaustive. With the advent of machine learning, new and perhaps more sophisticated approaches are being developed regularly. Many of these techniques have been developed to improve the ability of an algorithm to classify facial images into the correct groups (e.g., female or male), with increasing levels of accuracy for example, (Abbas et al. 2018). However, as the complexity of the mathematics behind these techniques increases, the difficulty in understanding the true differences between the subjects increases (Montavon et al. 2018). The analyses that have been explored in this thesis are more traditional in nature than machine learning algorithms and were chosen due to their comparability with mPCA. They serve to highlight the advantages and disadvantages of mPCA from the standpoint of a clinician who seeks to understand the differences in facial shape in a manner, rather than the ability of an algorithm to classify facial shapes correctly.

2.10 Overview

The first part of this thesis will investigate the claimed advantages of mPCA using 21 landmarks to simplify the initial analyses. A comparison of the descriptive and inferential results possible using the raw landmarks, conventional PCA, DFA and mPCA are presented using the influence of geographical location and sex on facial shape as an example. Once the advantages of mPCA are determined, the analysis is expanded to include 1000 quasi-landmarks. mPCA is utilised on these quasi-landmarks to determine the influence of sex, height, BMI, age, pubertal stage, metabolic factors, breathing disorders, maternal smoking during pregnancy and maternal alcohol consumption during pregnancy on the facial shape of English adolescents. Finally, the use of mPCA as a variable selection tool prior to PLSR is explored using 7160 quasi-landmarks, with the influence of the above variables on adolescent facial shape investigated in further detail.

3 METHODOLOGY

The methodology used in each of the three studies is outlined in here. As each study used different landmark strategies and analyses, an overview of the relevant methodology is also provided at the start of each study.

3.1 ETHICAL APPROVAL AND FUNDING

Ethical approval for the Croatian cohort was provided by University of Rijeka, Faculty of Medicine, number 2170-24-01-15-2, class 003-08/15-01/08, dated February 9th, 2015. Ethical approval for the use of the English sample was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees (B3166). Consent for samples was collected in accordance with the Human Tissue Act 2004. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Approval for the Welsh cohort was provided by the Southeast Wales Local Research Ethics Committee (04/WSE03/109). Approval for the Finnish cohort was provided by Oulu City Health Authority (4428/2006).

Although the work in this thesis was not specifically funded, The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website. The variables used in this thesis were specifically funded by: Wellcome Trust and MRC (core) 076467/Z/05/Z, University of Cardiff and NIH R01 DK077659.

3.2 POPULATIONS

3.2.1.1 Croatian sample

This convenience sample was obtained from a student population at University of Rijeka, Croatia in 2015 (n=73). The Croatian sample was older than the other populations but was assessed to determine the ability of mPCA to separate this group from the others. To determine the influence of geographical location and sex in a less biased manner, the mPCA analyses were also repeated without the Croatian sample.

3.2.1.2 English sample

Subjects were recruited through the ALSPAC cohort. The mothers were enrolled from the Avon area of Southwest, England. All mothers were required to be resident in the area and expected to give birth between 1st April 1991 and 31st December 1992. The initial enrolment included 14,541 pregnancies, who all completed at least one questionnaire or attended at least one clinic by 19th July 1999. Of the original 14,676 foetuses, 14,062 resulted in live births and 13,988 survived the first year of life.

When the original cohort of children were approximately seven years old, further subjects were recruited who had not joined the study originally. The subjects met the original inclusion criteria, with information sometimes gained from obstetric notes. This increased the included subjects to 14,541. Overall, there have been four additional phases of recruitment with 913, 456, 262 and 195 subjects enrolled in each respectively as of 24 years old. Data is available for 14,901 subjects who were alive at one year old. In addition, 10% of the ALSPAC cohort were enrolled in the Children in Focus group, who attended clinics at the University of Bristol at multiple time points between four and 61 months old. These children were randomly chosen from the subjects born in the last six months of the enrolment period and included 1432 families who attended at least one clinic. Subjects whose mothers moved to another area, did not attend and were therefore lost to follow-up or those included in another development study in the area were excluded. Further information on the cohort is available from Boyd et al. (2013) and Fraser et al. (2013). At approximately 15 years old, 9985 subjects were invited to attend additional clinics. Of these, 5235 subjects attended the clinic for a 3D facial scan.

3.2.1.3 Welsh sample

This cohort was recruited as a convenience sample through two schools in Pontypridd, Wales, UK (n=50). The facial scans were collected as part of a longitudinal study. The facial scans of the subjects at 15 years old were used for this thesis.

3.2.1.4 Finnish sample

This cohort was also a convenience sample, recruited from a school in Oulu, Finland as part of a longitudinal study (n=47). The facial scans of the subjects at 15 years old were used for this thesis.

3.3 FACIAL SCAN ACQUISITION

3D facial scans were taken of the English, Welsh and Finnish subjects using Konica Minolta VI-900 laser scanners. Class I eye safe lasers were used which have been suggested as having an operating accuracy of 0.3-0.5mm (Zhurov et al. 2005; Zhurov et al. 2010). Details of the image capture process are provided by Toma (2014). The laser scanner was placed at a reproducible distance from the subjects with lighting standardised via lamps. The camera was calibrated prior to each scanning session.

Toma (2014) also describes the protocol for ensuring reproducibility of subject positioning. Subjects were positioned in natural head position by asking subjects to look at a symbol placed between the scanners. Reference marks standardised the position of the subjects relative to the laser scanners. Subjects were asked to relax their facial muscles and swallow prior to the scan commencing. The facial scans took approximately seven seconds. 3D images were taken of the Croatian subjects using 3dMDface system, 3dDM Inc., Atlanta, GA, USA). Both systems have similar accuracy rates of approximately 0.5mm (Kau et al. 2004; White et al. 2020).

3.4 FACIAL SCAN PROCESSING

Scans of each half of the face were processed and merged using an in-house algorithm in Rapidform 2006 (Zhurov et al. 2005). This process aims to smooth the images and merge the images taken from the two laser scanners into one 3D facial image. To facilitate landmarking, the facial shells were superimposed at mid-endocanthion. This has been suggested as a possible alternative to the stable structures suggested by Björk (1955) in relation to lateral cephalograms given the proximity of mid-endocanthion to the cribriform plate (Zhurov et al. 2010). They were subsequently aligned using a cylindrical template.

In all studies, Generalised Procrustes Analysis (GPA) was used. GPA minimises the sum of the squared distances between the landmarks (Klingenberg 2021). This eliminates differences due to size, rotation, and position, thus isolating the influence of shape (Klingenberg 2021). The process is described by Stegmann and Gomez (2002). First, one shape is selected. The remaining shapes are aligned to this by calculating the centroid of each shape, aligning the shape centroids, normalising the shapes, and rotating the shapes to align with the new approximate mean. This is repeated until no significant difference is found between the mean shapes. In study one, GPA was used to facilitate manual landmarking.

3.5 LANDMARKING

This thesis uses sequentially more sophisticated landmark strategies. In study one and as a comparison in study two, 21 manually plotted landmarks were used as an initial investigation to investigate the feasibility of mPCA as an analytical technique in facial shape research. The 21 facial landmarks used were originally defined by Farkas (1994), and were used to represent the main facial features (Figure 6). The landmarks were plotted manually in Rapidform 2006. Previously, the 21 landmarks have been found to be both reliable and reproducible (Toma et al. 2009). Each of the samples had previously been landmarked by different researchers. Each of these researchers underwent a calibration exercise which was externally controlled by a separate, experienced researcher. A minimum intra-observer reliability of 1mm was aimed for, but for a minority of landmarks, a reliability of greater than 1mm was achieved. For these reasons, the differences between the geographical locations should be interpreted with caution. Instead, the differences due to geographical location were included to emphasise the ability of mPCA to differentiate between the different groupings.

In study two, to increase the number of landmarks assessed, 1000 quasi-landmarks were used as a comparison to the 21 landmark models (Figure 7). Quasi-landmarks provide a method for comparing one face to another by using a template to determine the points to compare from one subject to another. However, quasi-landmarks often do not represent an obvious facial feature (e.g., the middle of the cheek). This contrasts with manually plotted landmarks where, in general, a clear clinical feature is represented (e.g., endocanthion). Unfortunately, it was not possible to use the all of

the quasi-landmarks available (7160 quasi-landmarks) in the mPCA models due to restrictions on matrix sizes in MATLAB. A downsampling algorithm written by H. Matthews from KU Leuven was therefore used in MATLAB R2017b to reduce the 7160 landmarks to 1000 landmarks for the mPCA models (Figure 7).

In study three, the full 7160 three-dimensional quasi-landmarks were therefore used to provide as much detail as possible within the analyses.

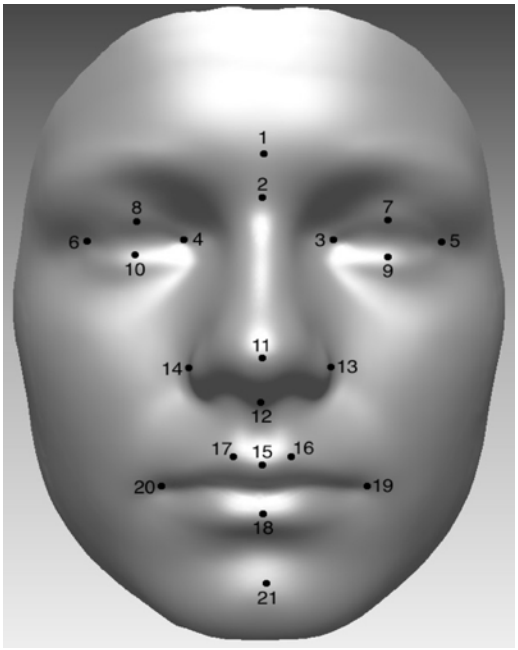


Figure 6: The 21 facial landmarks used to describe the main facial features as per the Farkas (1994) definitions: 1) Glabella; 2) Nasion; 3) Endocanthion (left); 4) Endocanthion (right); 5) Exocanthion (left); 6) Exocanthion (right); 7) Palpebrale superius (left); 8) Palpebrale superius (right); 9) Palpebrale inferius (left); 10) Palpebrale inferius (right); 11) Pronasale; 12) Subnasale; 13) Alare (left); 14) Alare (right); 15) Labiale superius; 16) Crista philtri (left); 17) Crista philtri (right); 18) Labiale inferius; 19) Cheilion (left); 20) Cheilion (right); 21) Pogonion

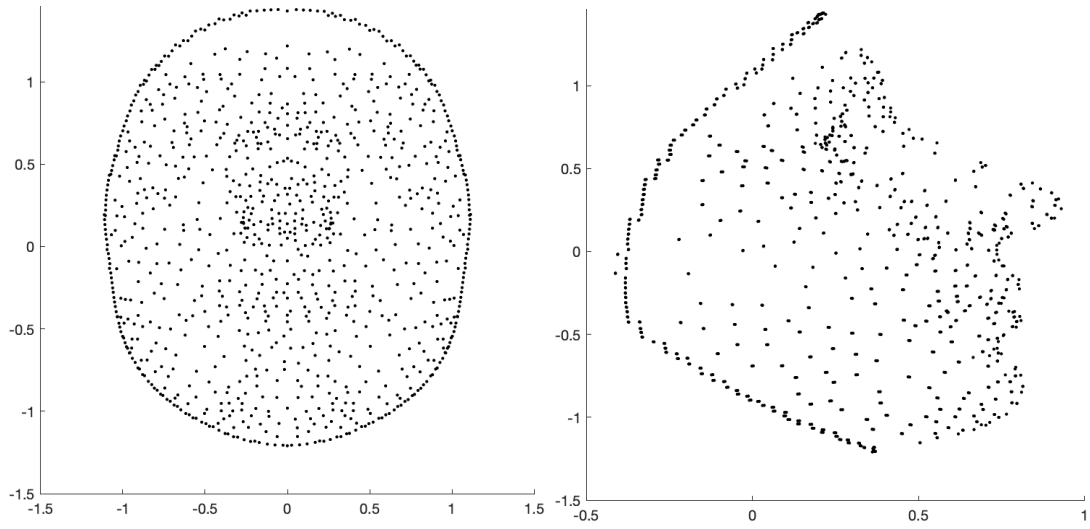


Figure 7: Quasi-landmarks retained after downsampling to 1000 landmarks. (a) Front view, (b) Profile

The quasi-landmarks were generated using the Meshmonk toolbox in MATLAB R2017b, developed in KU Leuven (available at https://github.com/TheWebMonks/meshmonk). The process is complex and aims to provide a facial mesh of 1000s of landmarks automatically, using algorithms to align the target face (subject) with a template face. The first step rigidly aligns the template face using five manually placed landmarks (endocanthion right and left, pronasale and cheilion right and left). A further rigid procedure follows, then a non-rigid iterative procedure is carried out to allow the template to deform gradually until it represents the target face as accurately as is possible (deterministic annealing). This process is explained in Claes et al. (2012), Snyders et al. (2014) and White et al. (2019). Snyder et al. (2014) document the following processes: one-to-one correspondences, transformation, outlier detection, multi-scale approach, deterministic annealing, symmetric registration, coherent point drift and image-based surface registration.

The 7160 quasi-landmarks were subsequently aligned using GPA, with a scaling procedure used during the process to remove the influence of size. In total, 7160 landmarks were generated in three dimensions (21480 datapoints).

3.6 VARIABLES

In study one, geographical location (Croatia, England, Wales and Finland) and sex were assessed (female and male). Sex was defined as sex at birth. These categorical variables were used to assess the feasibility of mPCA.

In studies two and three, the analyses focused on the ALSPAC cohort in isolation, with increased sample sizes. The ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: (http://www.bristol.ac.uk/alspac/researchers/our-data/). The variables assessed in this thesis have been suggested as having the potential to influence facial shape in previous research.

Information on the sex of the subjects was readily available from the survey data. Height and weight were measured with the subjects without shoes on and in thin clothing (Lawlor et al. 2010). Height and weight were subsequently used to calculate the BMI (BMI = weight (kg) / height^2 (m)). The pubertal stage was self-declared using Tanner Staging. This involved showing the subjects a picture scale of pubertal hair development and genital development and asking them to advise which picture they felt represented their stage of development (Emmanuel and Bokor 2020).

A fasting blood sample was taken at a recall clinic when the subjects were 15-16 years old. Subjects were asked to fast for a minimum of six hours, with those attending in the morning asked to fast overnight. The details of how the blood specimens were processed are available in Lawlor et al. (2010). After collection, the blood samples were spun and frozen immediately. The levels of metabolic factors were measured six to nine months after this. Cholesterol, triglycerides and HDL were determined using enzymatic reagents. LDL was determined using the Friedwald equation which considers cholesterol, HDL and triglycerides. Fasting insulin levels were determined using enzyme linked immunosorbent assay (ELISA) and glucose using an automated assay.

Asthma was determined via questionnaire with mothers reporting the presence of wheezing at 7.5 years old (Al Ali et al. 2014a). Asthma between 0-3.5 years old was

also determined via questionnaire and again defined as wheezing. Mothers were asked to confirm if wheezing was present at six months and 42 months. Atopy was determined via skin prick test on the left forearm at 7.5 years old to six common allergies (peanut, cat hair, milk, grass, mixed nuts and dust mites). Positive and negative controls were used (1% histamine solution and dilutant respectively). A positive test was determined after 10 minutes where the weal diameter was equal to or exceeding 1mm (Henderson et al. 2008; Al Ali et al. 2014a).

Maternal smoking levels before pregnancy, during the 1st trimester and the 2nd trimester were self-reported via questionnaire at 18 weeks gestation. The mothers were asked to declare their smoking levels on an ordinal scale: 0 cigs/day, 1-4 cigs/day, 5-9 cigs/day, 10-14 cigs/day, 15-19 cigs/day, 20-25 cigs/day and 30+ cigs/day. Maternal alcohol consumption was also determined via questionnaire at 18 weeks gestation. The mothers were again asked to declare this on an ordinal scale: 0 glasses/week, less than 1 glass/week, 1+ glasses/week, 1-2 glasses/day, 3-9 glasses/day and 10+ glasses/day). They were asked to class a glass of alcohol as "one unit of spirits, half a pint of beer or cider or a glass of wine". Alcohol consumption after the baby first moved was used as a proxy for the 2nd trimester.

3.7 DATA CLEANING AND EXPLORATION

3.7.1 Exclusions

For the English sample, a total of 399 subjects were excluded due to not having their face scanned at the clinic, facial scan of poor quality, not of Caucasian descent or facial dysmorphology present. Facial scans were not pre-processed if they were not of sufficient quality after registration of the right and left facial shells and were therefore removed from the analyses (n=89). This resulted in facial scans being available for 4747 children (2233 males and 2514 females) (Toma 2014). Subjects were further excluded if they were a twin or triplet, with one subject from each family group retained and/or if there was missing data on any of the variables included in the separate models (n=3301). The latter decision was made to keep the mPCA models comparable. An alternative method would have been to include all possible subjects

in each model but the results of each would not have been directly comparable as each model would have included different numbers of subjects.

Bookstein (2019) suggests that a group should not have less than ten subjects for between-groups PCA. This was also used as a benchmark for mPCA, however, where possible, the number of subjects per group was kept as large as possible. As four subjects self-declared a Tanner stage of one with regards to pubic hair and eight subjects self-declared their development stage of two, these subjects were excluded to maintain group sizes greater than ten. To keep the pubertal stage analysis balanced, these groups were also removed for the Tanner stage with regards to genital development despite 14 subjects self-declaring a Tanner stage of two. A final cohort of 1411 subjects remained.

3.7.2 Distribution of landmark data

It has been suggested that DFA requires normal distribution of the data set (Moore, 2013). The distribution required for PCA is less clear. Some authors have suggested PCA requires normally distributed data, however, Jollifee (2002) disagrees. To avoid issues, the distribution of the scaled landmark data (21 landmarks) was visualised using KS density plots in MATLAB R2017b and confirmed to be normally distributed (Figures 8-10). These plots were used as an alternative to histograms to aid visualisation of multiple landmarks on one plot.

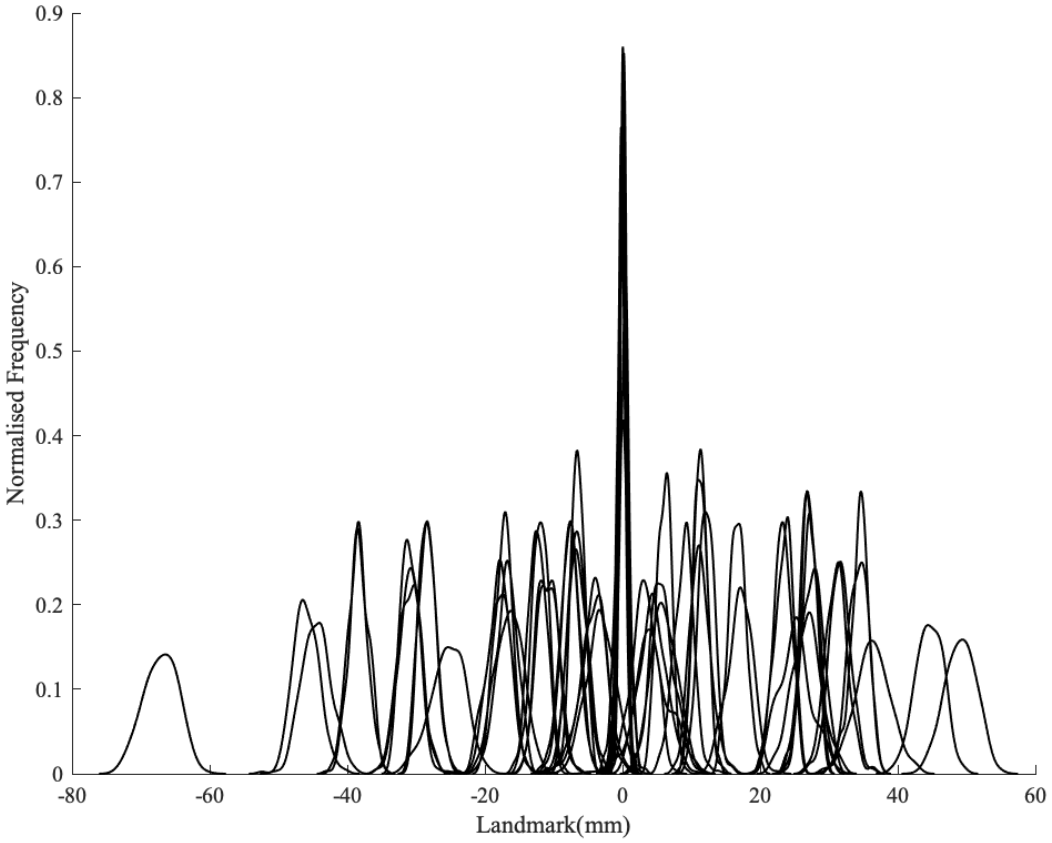


Figure 8: KS density plot demonstrating the normal distribution of the landmarks in the subjects included in the four European population analysis (n=249).

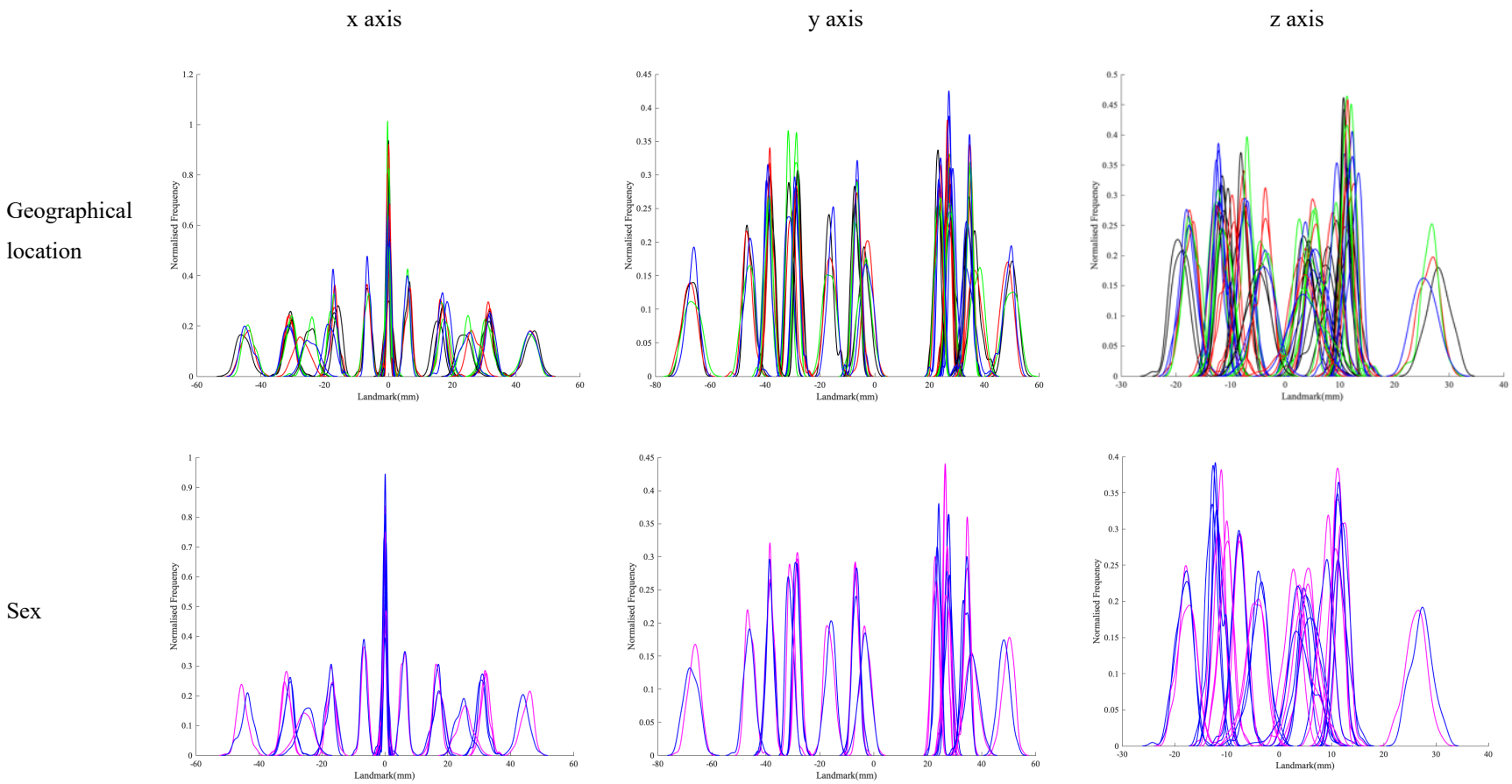


Figure 9: KS density plots demonstrating the normal distribution of the landmarks in three dimensions for the four European population analysis (n=249). Grouped by population and sex.

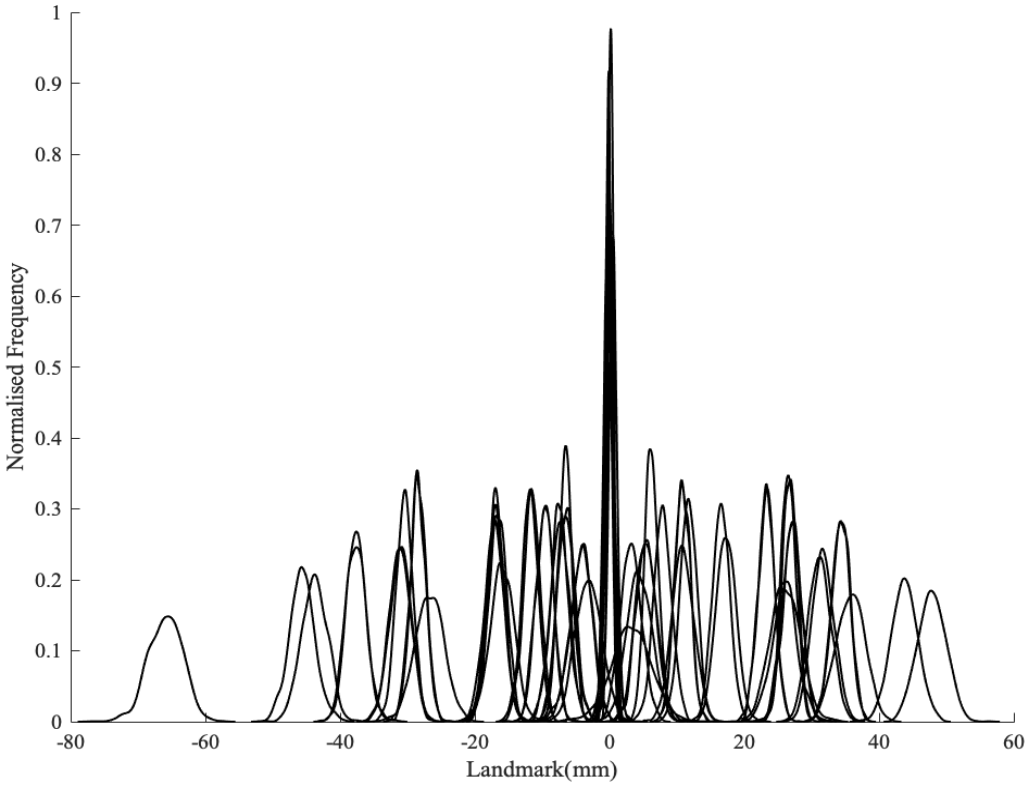


Figure 10: The 21 landmarks were normally distributed in the ALSPAC cohort and visualised via KS density plots (n=1411).

3.7.3 Outliers

3.7.3.1 Outliers in landmark data

As the landmarks were placed manually, outliers may be due to researcher error rather than true anatomical variation. Outlying landmarks were detected using the MATLAB function "isoutlier" at both 3SD and 4SD thresholds. For study one, one outlier (English female) was detected at 4SD in the x-axis at pronasale (Figure 11). This subject was removed from the analyses.

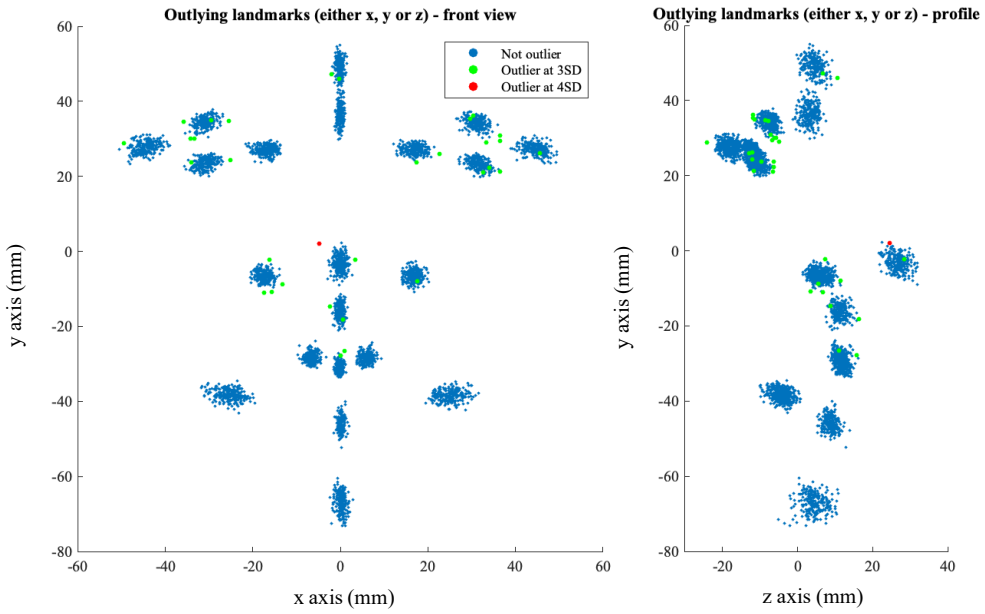


Figure 11: Scatter plots highlighting outlying landmarks at 3SD (green) and 4SD (red) for the four European populations. One landmark, pronasale in the x-axis, associated with an English female, was an outlier at 4SD. The subject was removed from the analyses.

Outlying landmarks were also found at a 4SD level in the ALSPAC specific analyses (Figure 12). However, following visual inspection of the faces and the data set as a whole prior to exclusions due to missing data, it was decided to include these subjects in the analyses. As only one researcher placed the landmarks for the ALSPAC specific analyses, who had undergone a calibration as part of this process, further reassurance was gained with regards to including these subjects. No outliers were detected in the 1000 quasi-landmark data using the "isoutlier" function at a 4SD threshold.

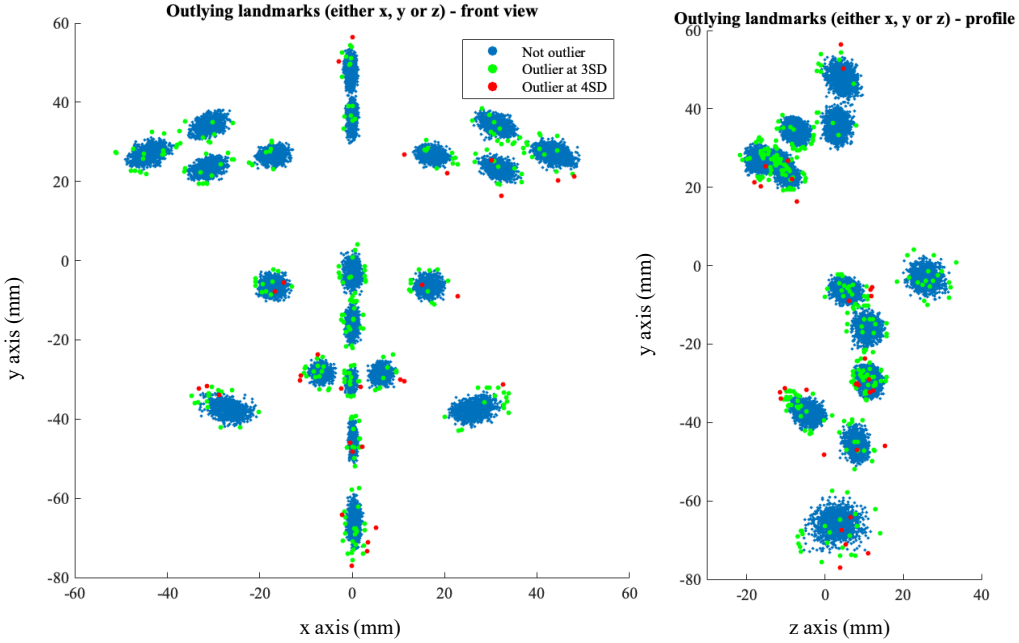
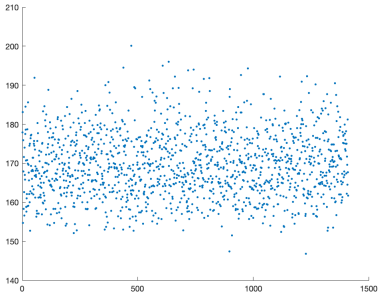


Figure 12: Scatter plots highlighting outlying landmarks at 3SD (green) and 4SD (red) for the ALSPAC specific analyses. Several landmarks were found to be outliers, however, following inspection of the faces and the data set as a whole prior to exclusions, it was decided to include these subjects in the analyses.

3.7.3.2 Outliers in variable data

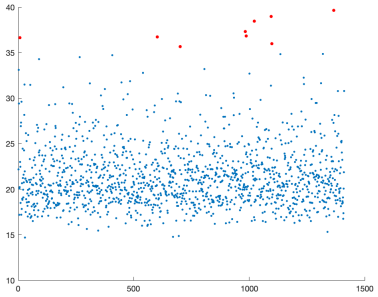
mPCA can deal with outliers in the variables using discretisation (any subjects with very high or very low levels of a variable will be grouped with those closest to them). There was therefore no requirement to remove any subjects from the mPCA analyses on this basis. However, groups with small sample sizes required removing or combining with other groups. By contrast, regression methods can be sensitive to outliers (Tabachnick and Fidell 1996). The PLSR models in study three were therefore run with and without outliers to investigate the impact of their inclusion. This was carried out using the "isoutlier" function in MATLAB R2017b with a threshold of 4SD from the mean. No outliers were found with regards to height. Nine subjects were found to be outliers with regards to BMI. All these subjects were in the obese or severely obese range. The BMIs of these subjects are not out-with potentially real values. Six outliers were found with regards to insulin. Two of these values were extreme outliers. Two outliers were found for cholesterol and LDL, whilst five outliers were found for triglycerides and very low density lipids (VLDL) (Figure 13).

Height



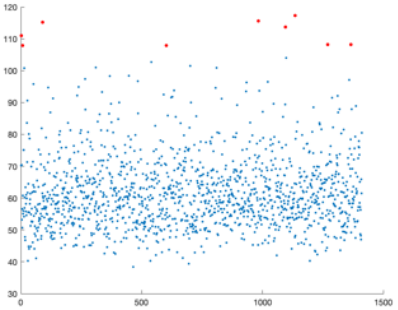
Number of outliers = 0

BMI



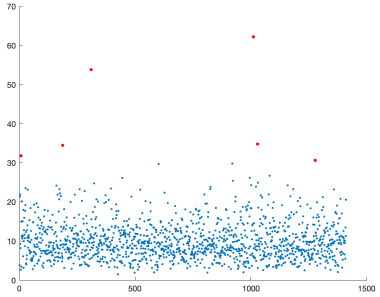
Number of outliers = 9

Weight



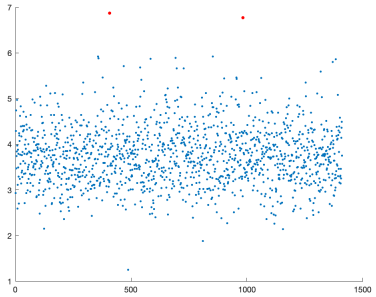
Number of outliers = 9

Insulin



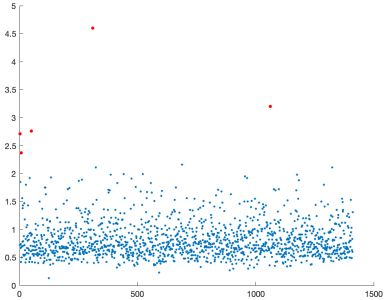
Number of outliers = 6

Cholesterol



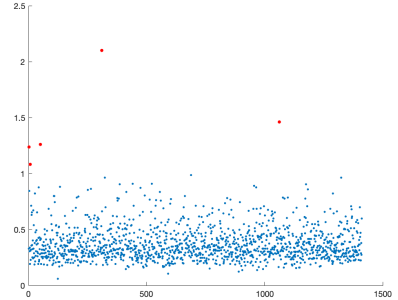
Number of outliers = 2

Triglycerides



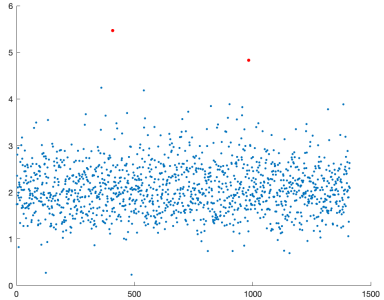
Number of outliers = 5

VLDL



Number of outliers = 5

LDL



Number of outliers = 2

Figure 13: Outliers detected at 4SD from the mean for the variables included in the PLSR models.

3.7.4 Correlations

3.7.4.1 Correlations in landmark data

To investigate the correlation between the landmark data used in study one, a correlation matrix using Pearson's correlation was visualised using MATLAB 2017b. The correlation matrix confirms that many of the landmarks are moderately to strongly correlated (Figure 14). This is to be expected as facial features all have a similar shape and are relatively predicably positioned in relation to each other on the face. This confirms that multicollinearity is present in the landmark data.

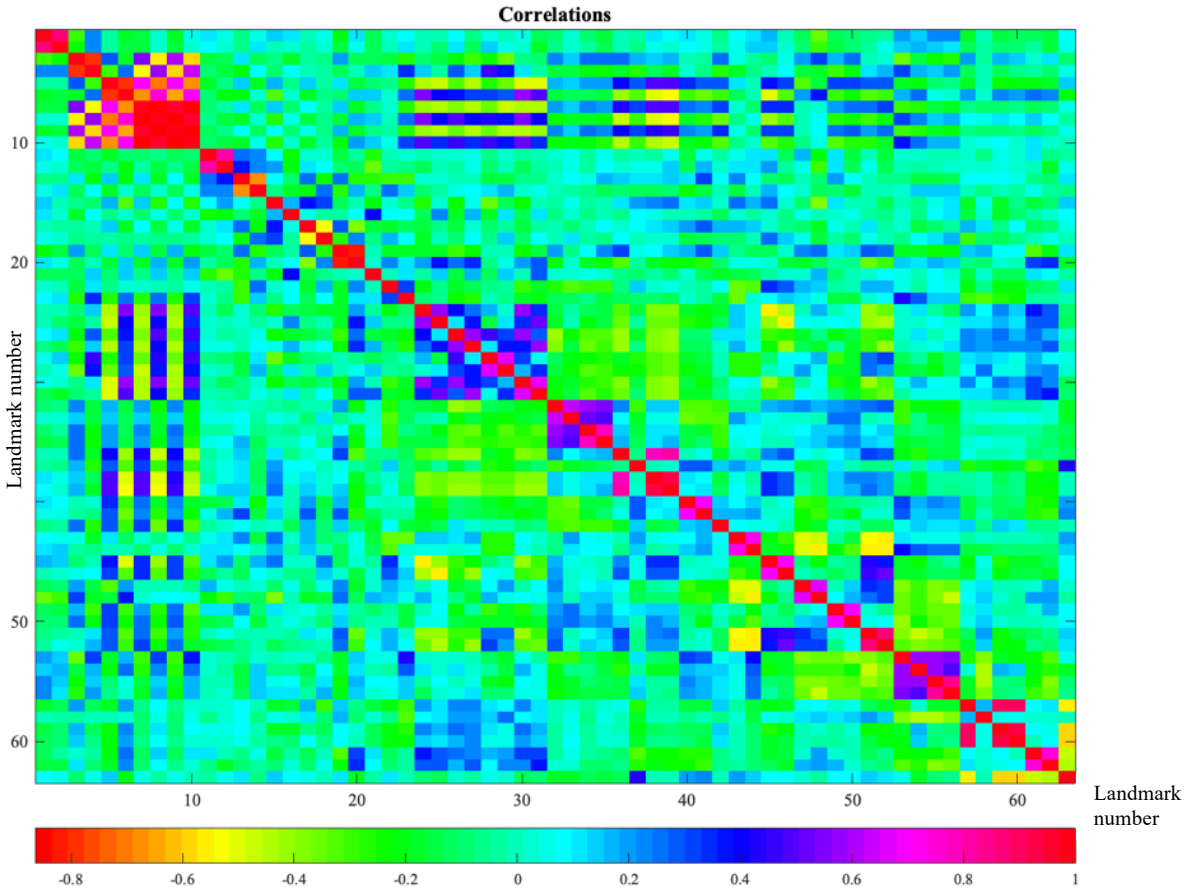


Figure 14: Matrix of Pearson correlation coefficients showing the landmarks that are strongly correlated (red) and weakly correlated (more light blue). As would be expected, the highest correlations are grouped by facial feature.

3.7.4.2 Correlation between variables

Assessment of the correlation between the variables used in the ALSPAC specific studies (studies two and three) provided information on which variables were more likely to act as confounding factors in each of the mPCA models. The correlations were visualised using scatter plots, box plots and heat plots. The test used depended on the type and distribution of the data (summarised in Table 4). Scatter plots were constructed as a matrix in SPSS v25 to visualise the correlation between all the continuous variables. As it is difficult to ascertain patterns in the scatter plots for the dichotomous and ordinal data, box plots were generated in MATLAB R2020a to visualise the association between categorical and continuous variables. These were collated using this tiled layout function. Finally, heat/deviation maps were generated in MATLAB R2017b to visualise the original and constructed categorical variables. These were also collated as a matrix by hand. There is no option for calculating the rank biserial correlation in SPSS. Thus, the mean ranks were calculated for each group using the Mann Whitney U test and the following formula used, as suggested by Gray and Kinnear (2012):

$$r_g = \frac{2(M_1 - M_2)}{n_1 + n_2}$$

Where, M_1 = mean rank of group 1, M_2 = mean rank of group 2, n_1 = total subjects in group 1 and n_2 = total subjects in group 2.

Table 4: Correlation tests used

	Dichotomous	Ordinal	Continuous (normally distributed)	Continuous (skewed)
Dichotomous	Phi	Cramer's V	Point Biserial	Rank Biserial
Ordinal	Cramer's V	Spearman's Rank	Spearman's Rank	Spearman's Rank
Continuous (normally distributed)	Point Biserial	Spearman's Rank	Pearson's	Spearman's Rank
Continuous (skewed)	Rank Biserial	Spearman's Rank	Spearman's Rank	Spearman's Rank

The correlations between the variables are presented in Figures 15-17. Height is moderately strongly correlated with sex, which is to be expected, as after puberty males are in general 15cm taller than females (Desai 2000). The strength of this correlation is maintained following categorisation of the height variable. Both measures of pubertal stage are weakly correlated. With regards to the metabolic factors, insulin is weakly correlated with BMI, with a similar magnitude seen when considering both the continuous variables and constructed ordinal variables. Cholesterol is highly correlated with LDL and weakly correlated with VLDL and triglycerides. This is true when assessing the variables in their original continuous form and once categorisation was carried out. In turn, triglyceride levels were highly correlated with VLDL, and this is again maintained after categorisation. This is to be expected as the function of VLDL is to carry triglycerides to adipose tissue and muscle (Freeman and Walford 2016). A weak correlation is also seen between cholesterol and HDL, and triglycerides and LDL. In general, the correlation is less strong when assessing the categorised variables in comparison to the same variables prior to categorisation. However, the strength of the correlation falls into the same broad description for all variables, except for VLDL and LDL, which fall from weakly correlated (0.336) to minimal correlation (0.274) after categorisation. Asthma in early childhood is weakly correlated with asthma at 7.5 years old. It may have been expected that this correlation would be higher but indicates that asthma may be diagnosed after 3.5 years old, or that those diagnosed with asthma in early childhood recover by 7.5 years old. Atopy and hay fever are also weakly correlated. It may have been expected that this correlation would have been stronger as hay fever is used in the diagnosis of atopy. However, as other factors are involved in an atopy diagnosis, these must have a strong influence on the diagnosis without the presence of asthma. Interestingly, smoking and alcohol consumption do not appear to be strongly correlated. Alcohol consumption in the 1st trimester is weakly correlated with alcohol consumption before pregnancy. This is perhaps to be expected, particularly as many mothers do not know that they are pregnant in the early stages of pregnancy. The correlation between alcohol consumption during the 1st and 2nd trimester is stronger. Smoking before pregnancy highly correlated with smoking during the 1st trimester and continuing to smoke into the 2nd trimester.

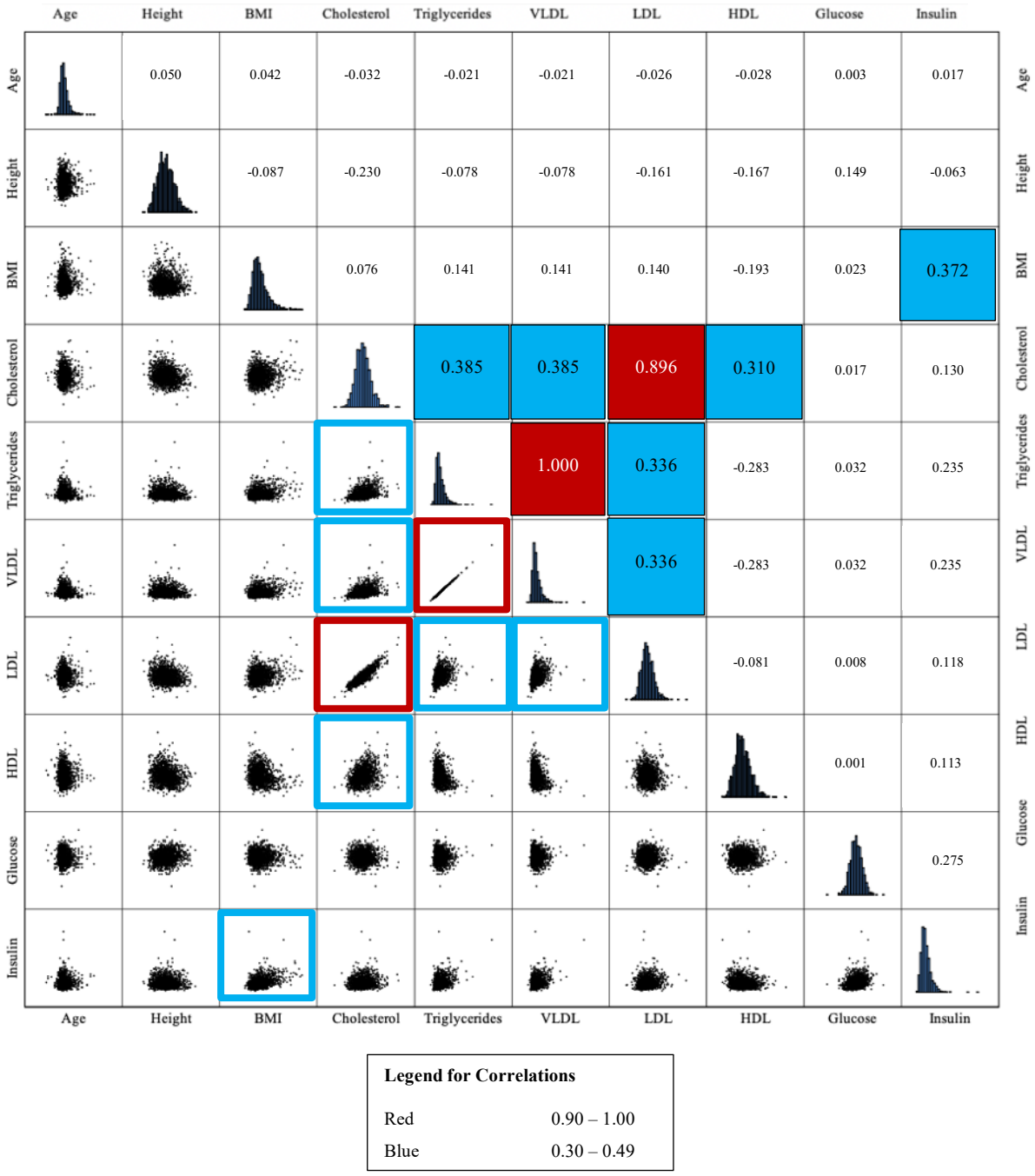


Figure 15: Correlation between the continuous variables. On the left of the diagonal are scatter plots, displayed as a matrix, demonstrating the relationship between the continuous variables prior to categorisation. The distribution of each variable is displayed as a histogram on the diagonal. On the right of the diagonal are the correlation coefficients.

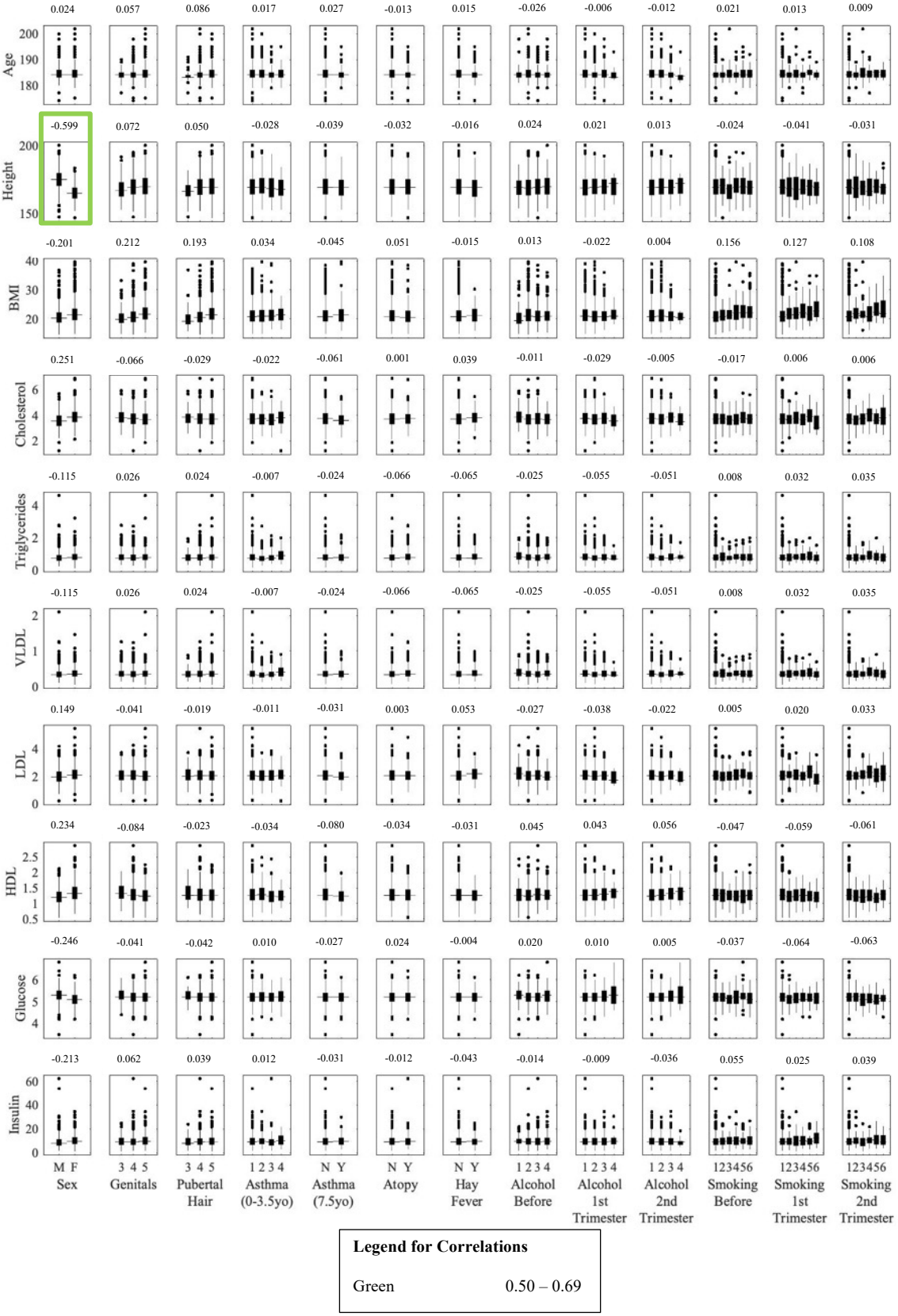
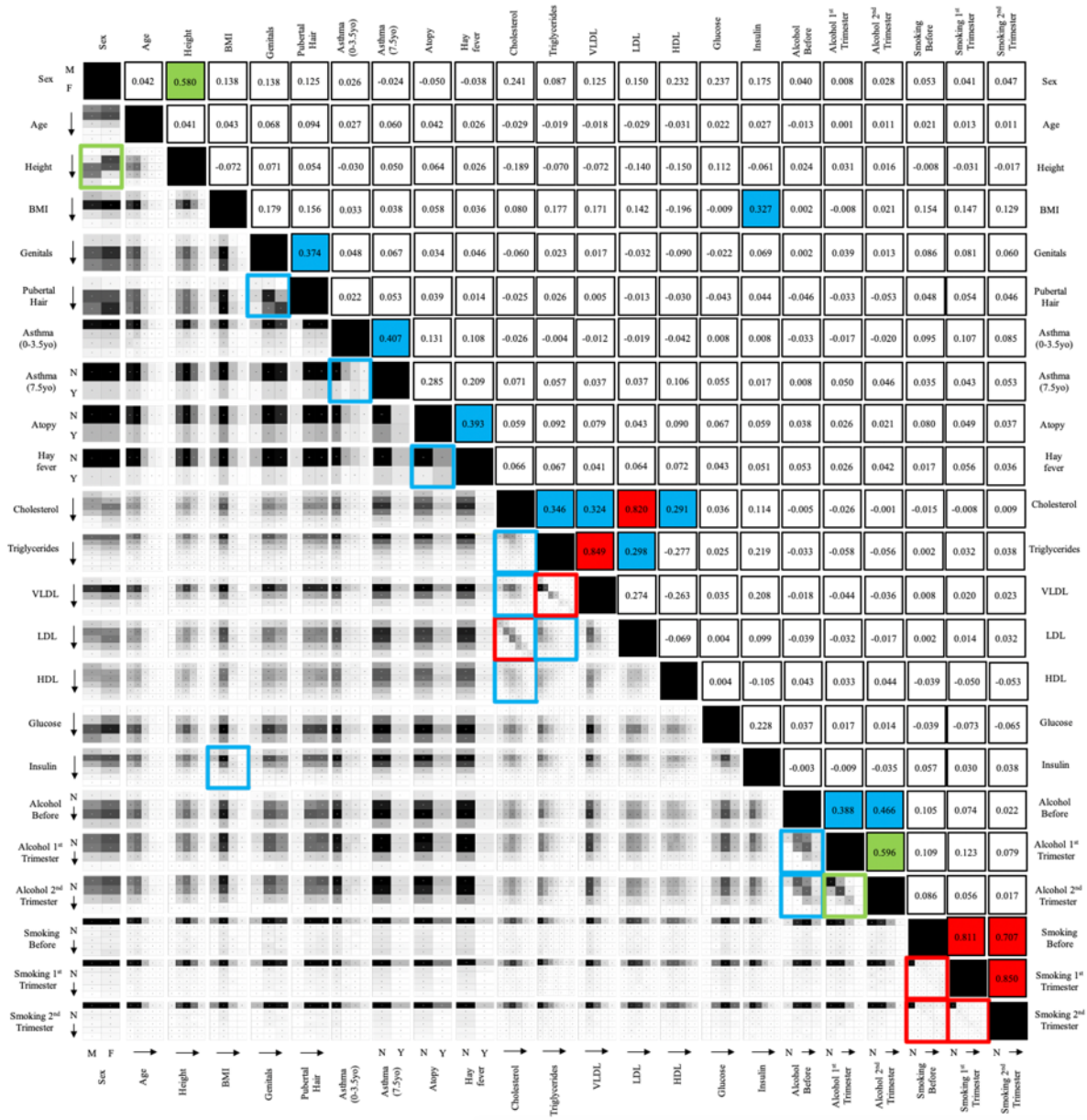


Figure 16: Box plots displaying the correlation between the categorical and continuous variables. The correlation coefficients are presented above each box plot.



Legend for Correlations	
Red	0.90 – 1.00
Green	0.50 – 0.69
Blue	0.30 – 0.49

Figure 17: Correlation between the groups after categorisation. On the left of the diagonal are heat plots, displayed as a matrix, indicating the correlation between the groups. White = 0 subjects, black >= 500 subjects. On the right of the diagonal are the correlation coefficients.

3.7.4.3 Standardisation

Standardisation has been suggested as a useful pre-processing stage prior to conventional PCA (Baxter 1995). Standardisation involves subtracting the grand mean and dividing by the standard deviation to bring each of the variables onto the same scale. In this thesis, all the landmarks were recorded in millimetres and were therefore on a similar scale. After exploring the influence of standardisation in mPCA in Appendix 1, on balance, it was decided not to standardise the landmark data prior to mPCA to make the results more easily interpretable. However, the groups of the standardised landmarks were separated better than the non-standardised landmarks.

3.7.5 Discretisation

Multilevel PCA requires discretisation (categorisation) of continuous data. Where well established categories were available (BMI, pubertal stage) these were used. Where no established categories were existing, groups were constructed by initially separating subjects into healthy and pathological groups, where these distinctions were available. Subjects within normal/healthy range were subsequently categorised so that the groups were of equal width, with a reasonable number of subjects per group. This allowed the discretisation process to be flexible, to gain as much meaningful information from the data whilst maintaining groups with reasonable sample sizes. However, this approach introduced a level of subjectivity.

3.7.6 Final sample sizes

The final sample sizes are provided at the start of each of the relevant results chapters.

3.7.7 Statistical analyses

An overview of MANOVA, conventional PCA, DFA, mPCA and PLSR can be found in 2.9 Analysis of 3D facial landmarks (p.40) with further detail in the results chapters.

- Study one (21 landmarks): Landmarks only, conventional PCA to facilitate MANOVA, DFA and mPCA (3-level)
- Study two (1000 quasi-landmarks): mPCA (2-level)
- Study three (7160 quasi-landmarks): PLSR with variables explained 2% of the variation in their mPCA models.

4 STUDY 1: INITIAL EXPLORATION OF MPCA IN FACIAL SHAPE RESEARCH USING A THREE-LEVEL MODEL (GEOGRAPHICAL LOCATION, SEX AND WITHIN-GROUP VARIATION)

4.1 INTRODUCTION

As discussed previously, it has been claimed that mPCA is a useful analytical tool for facial shape (Farnell et al. 2017). The mPCA model was expanded to a three-level model here. Prior to utilising mPCA to determine the influence of numerous variables on facial shape, the relative advantages and disadvantages of mPCA are explored in this study. These were demonstrated by comparing the results possible with mPCA to those possible using raw landmarks only, conventional PCA and DFA. Twenty-one landmarks were used to allow clear visualisation of the differences and to keep analyses manageable in this initial phase of exploration. Geographical location and sex were investigated as these variables were anticipated to explain a relatively large proportion of facial shape variation, thus provided a sensible starting point for comparing the analytical techniques.

4.2 AIMS

- To explore the suitability of mPCA as an analytical tool for categorical data (similar sample sizes in each group) compared to raw landmarks, conventional PCA and DFA.
- To determine the influence of geographical location and sex on the facial shape of four European populations.

4.3 NULL HYPOTHESES

- Geographical location is not associated with the facial shape of the subjects from the Croatian, English, Welsh and Finnish populations.
- Sex is not associated with the facial shape of the subjects from the Croatian, English, Welsh and Finnish populations.

4.4 METHODOLOGY OVERVIEW

An overview of the methodology used in this study is provided below.

4.4.1 3D facial scan acquisition, processing and landmarks

The 3D facial scans were acquired and processed as documented previously. GPA was conducted on the 3D facial scans to facilitate landmarking. Scaling of the images removed the influence of size, thus isolating the influence of shape (3.4. Facial scan processing, p; 53). Twenty-one landmarks were placed manually using Rapidform 2006 (3.5 Landmarking, p. 52). The researchers who placed the landmarks underwent a calibration exercise.

4.4.2 Variables

Geographical location and sex were modelled explicitly in this section. Twenty-one manually plotted landmarks were used as outcomes variables as a proxy for the main facial features.

4.4.3 Data cleaning and exploration

Subjects were excluded if they had obvious craniofacial dysmorphology or their 3D facial scan was of insufficient quality. The landmarks were normally distributed (3.7.2 Distribution of landmark data, p. 59), with multicollinearity present in the landmark data (3.7.4.1 Correlations in the landmarks, p. 65). One outlier was detected at pronasale. This subject was removed from the analyses (3.7.3.1 Outliers in landmark data, p. 62).

4.4.4 Final sample

The samples are detailed in 3.2 Populations (p. 51). In total, 249 subjects were included following exclusion of one outlier in the English sample. Subgroup sample sizes are available in Table 5. Sample size calculations were not possible due to the novelty of the methodology.

Table 5: Subject demographics for investigation of the influence of geographical location and sex on the facial shape of four European populations

Geographical location	Sex	Age (± SD)	BMI (± SD)
Croatia (n = 73)	Female (n = 38)	22.98 ± 2.50	20.67 ± 1.78
	Male (n = 35)	23.11 ± 3.08	23.85 ± 2.42
England (n = 79)	Female (n = 39)	16.38 ± 0.29	22.08 ± 3.75
	Male (n = 40)	16.22 ± 0.18	22.42 ± 3.50
Wales (n = 50)	Female (n = 23)	16.15 ± 0.32	23.86 ± 2.69
	Male (n = 27)	16.11 ± 0.31	23.65 ± 4.66
Finland (n = 47)	Female (n = 23)	16.18 ± 0.49	20.68 ± 2.48
	Male (n = 24)	16.44 ± 0.37	20.94 ± 2.49

4.4.5 Analyses

4.4.5.1 Landmarks only

Using MATLAB R2017b, the 21 landmarks for each of the subjects were plotted as scatter plots (*x*- against *y*- for a front view and *z*- against *y*- for a profile view). The mean of landmarks for each group were also visualised using MATLAB 2017b. The differences were visualised using scatter plots (*x*- against *y*- for a front view and *z*- against *y*- for a profile view). The standard deviations were calculated in Excel v16.37 and visualised as tables. MANOVA was not carried out due to multicollinearity in the landmark data and the number of landmarks exceeding the number of subjects per group in the geographical location analysis (Tabachnick and Fidell 1996).

4.4.5.2 Conventional Principal Component Analysis

Conventional PCA is explained in 2.9 Analysis of 3D facial landmarks (p. 40). Conventional PCA was carried out for two purposes. Firstly, to provide a comparison to mPCA and explore the potential benefits of mPCA in the data exploration phase. The second purpose was to carry out dimension reduction to facilitate MANOVA. Conventional PCA was conducted on the raw landmarks using the "pca" function in MATLAB R2017b.

After inspection of the eigenvalue plot, 15 PCs were explored in further detail. Many methods have been proposed to determine how many PCs to discard. In this thesis, the method was based on Cattell's scree test for factor analysis (Cattell 1966). The

eigenvalues of each of the PCs were visualised with the PCs discarded after the curve began to plateaux (Figure 18). This is an efficient, flexible method for choosing which eigenvalues to explore further. However, it is subjective.

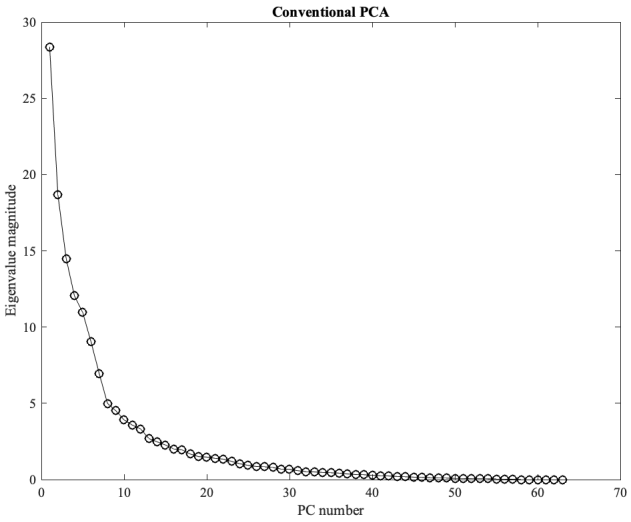


Figure 18: Eigenvalue plot showing the pattern of eigenvalue magnitudes. PC1 represents the most variation in the dataset. PC63 explains the least. The curve begins to plateaux at PC15.

The component scores were standardised by subtracting the grand mean and dividing by the standard deviation (square root of the respective eigenvalue). MANOVA was carried out on the standardised component scores in SPSS v25 to determine inference. Univariate ANOVAs were also carried out to determine how the PCs differed with geographical location. These were defined in the SPSS output as "tests of between-subjects effects" after implementation of MANOVA. Bonferroni correction was used to account for multiple testing, with significance set at 0.05/15 = 0.003. The standardised component scores that met the significance threshold were explored in further detail. As the scope of this analysis was primarily to provide a comparison with mPCA, post-hoc Tukey tests were not applied.

The difference in the mean standardised components scores for each grouping were visualised via scatter plots, generated in MATLAB R2017b. Colour coding was used to highlight differences due to geographical location and sex. The facial shape differences associated with each PC were also visualised via scatter plot. The square root of the respective eigenvalue and eigenvectors were added and subtracted to the

mean of the landmarks to determine the magnitude and direction of variation at each landmark.

4.4.5.3 Discriminant Function Analysis

DFA is explained in 2.9 Analysis of 3D facial landmarks (p. 40). A "rule of thumb" with regards to sample sizes for DFA is that the number of subjects per group should be at least four times the number of landmarks (Stella 2019). Others have suggested that the size of the smallest group should be greater than the number of outcome variables (Tabachnick and Fidell 1996). Given that the former is more conservative, it was used to demonstrate the possible limitations of DFA in the context of data sets with large numbers of outcome variables.

DFA was conducted in SPSS v25 on the standardised component scores rather than the landmark data given the sample size restriction a minimum of four times the subjects per group compared to the landmark data (Stella 2019). The scores were grouped by geographical location for the first analysis and sex for the second. Three discriminant functions (DFs) were possible for geographical location (four groups minus one) and one DF when assessing sex (two groups minus one). The DF scores were visualised using scatter plots in MATLAB R2017b. Inference was determined via Wilks Lambda in SPSS v25.

4.4.5.4 Multilevel Principal Compenent Analysis

mPCA is explained in 2.9 Analysis of 3D facial landmarks (p. 40). A three-level, non-nested mPCA model was developed utilising the median for the covariance matrices – geographical location (level one), sex (level two) and within-group variation (level three). The model structure is summarised in Figure 19. Although geographical location was placed a level one and sex at level two, as this is a non-nested model, this placement is interchangeable and non-hierarchical. The component scores were constrained within the model to a maximum and minimum of 3SD from the mean. The component scores were standardised by subtracting the grand mean of the component scores and dividing by the square root of the respective eigenvalue (the standard deviation).

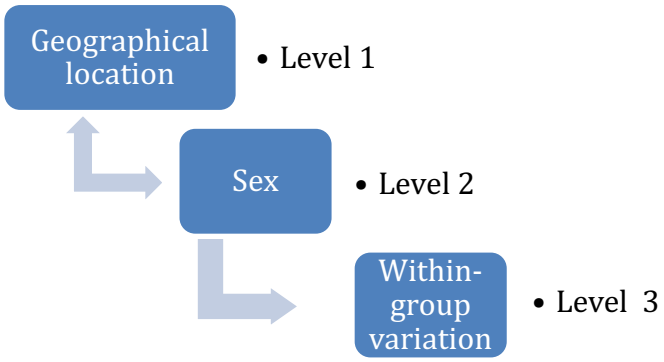


Figure 19: Three-level mPCA model for study one. There is no hierarchy between levels one and two in a non-nested model.

The component scores were calculated using all model levels simultaneously. This is possible using MATLAB's Global Optimizer to fit to the final model. 100,000 iterations were used to find the best solution for the component scores for each subject, at each level. The standardised component scores for each group were visualised as scatter plots in MATLAB R2017b to determine patterns in the data. The importance of each PC at each level (and therefore each variable) was expressed as a percentage of total variation:

$$Relative\ importance\ of\ variable\ (\%\ total\ variation) = (\Sigma(\text{Eigenvalues variable of interest}))/(\Sigma(\text{All eigenvalues in model})) \times 100$$

The facial differences explained by each PC were assessed by adding and subtracting the square root of the respective eigenvalue/ eigenvector to the mean facial shape of all the subjects. The differences were visualised as scatter plots in MATLAB R2017b.

MANOVA was used to determine inference at each level of the model. As MANOVA can be sensitive to outliers and to provide a robust test in the presence of non-normally distributed component scores, an in-house bootstrapping protocol was used in MATLAB R2017b to check the *p*-values. Component scores were generated via the "datasample" function in MATLAB R2017b which randomly subsamples the data with replacement. 1000 subsamples were generated via this procedure to find a *p*-value of precision 10^{-4}.

4.5 RESULTS

4.5.1 Influence of geographical location

4.5.1.1 Landmarks only

The differences in the landmark data, grouped by geographical location, can be visualised in Figure 20. The mean of each landmark and the associated standard deviations are summarised in Appendix 2. In the transverse direction (*x*-axis), there is very little variation in the midline. The Croatians are most differentiable at the orbits, presenting with increased orbital widths and therefore a reduced distance between the eyes. The difference between the Croatians and the other populations is between 2-3mm and with the biggest difference between the Croatians and the Finnish. The differences are consistent across the sexes. The English appear most differentiable at the corners of the mouth by as much as 4mm. This is consistent across both sexes; however, the English females have a wider mouth than the English males.

In a vertical direction (*y*-axis), the landmarks in the midline demonstrate larger variation in comparison to the *x*-axis. This is true between and within the groups. The largest difference is between the Finnish population and the others at nasion, with the Finnish population presenting with a more inferior nasion position by 3-4mm. The Finnish females have the most inferior nasion position.

In the anteroposterior direction (*z*-axis), the English and Finnish populations appear to be the subtly more class II, with a retrusive pogonion in comparison to the other populations by approximately 1mm. The Croatians appear to have the most prominent pronasale by 2-3mm. This is particularly true of the Croatian males. The corners of the eyes also appear more inset in the Croatians (1-2mm).

When visualising the overall mean, and therefore attempting to determine the broad differences between the groups, the Croatians are perhaps most distinguishable. However, as GPA centres the faces on the mean of all the faces, the overall mean for each population is almost zero. It is therefore challenging to summarise the overall differences between the groups as each landmark requires investigation.

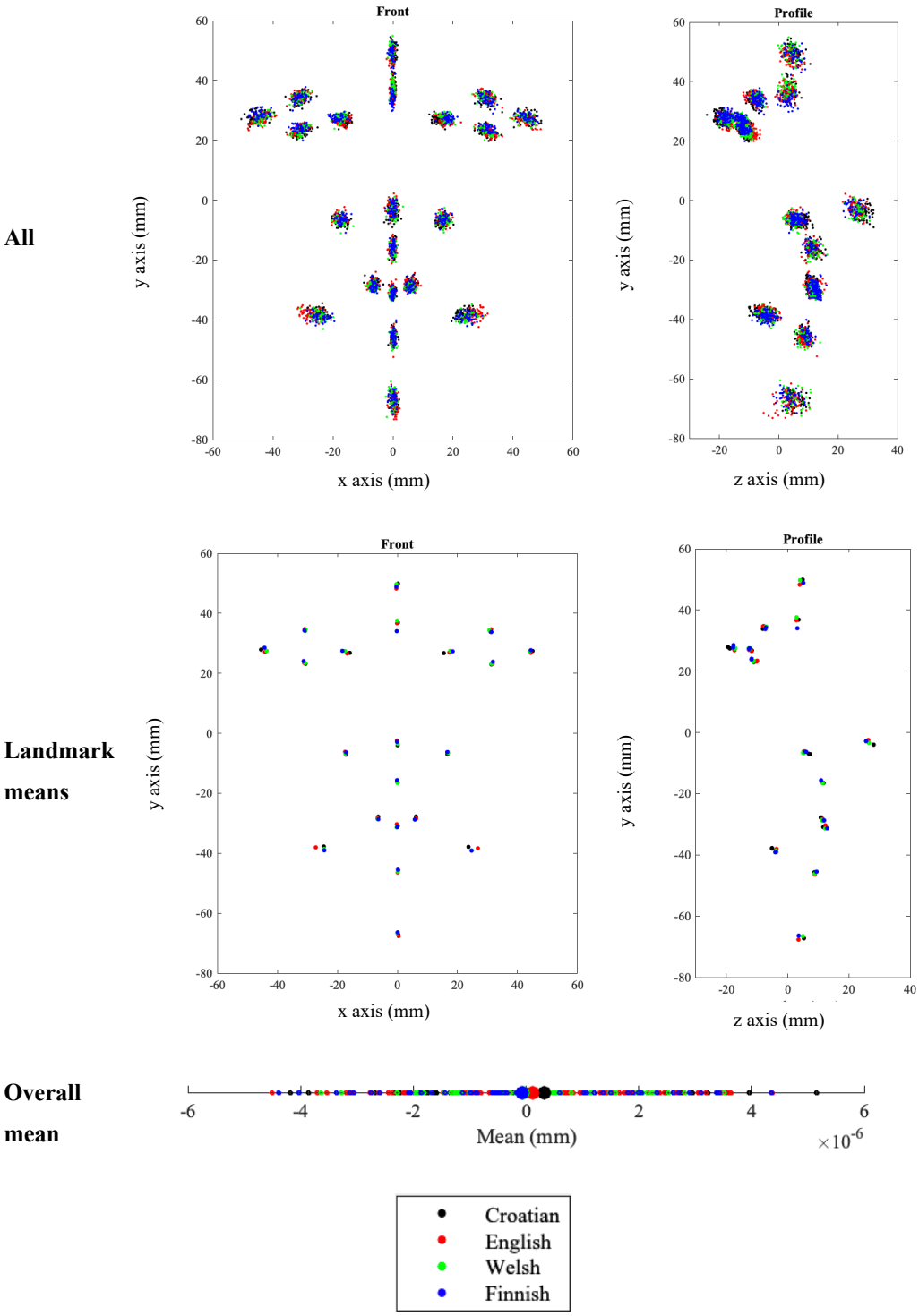


Figure 20: The 21 landmarks (facial meshes processed via GPA with scaling) grouped by population. Both between-group and within-group variation can be visualised. (a) All of the landmarks are visualised. (b) The landmark means are visualised. (c) The overall mean of the landmarks is visualised (the means are almost zero as the GPA scaling procedure centres the faces on the mean), thus does not help distinguish between the groups.

4.5.1.2 Conventional Principal Component Analysis

The MANOVA model was significant at a 0.05 level. Of the 15 standardised PCs assessed in further detail, PC2-10 were found to be significant at a 0.003 level with regards to geographical location using univariate ANOVAs (Table 6). The Croatians were clearly separated from the other populations at PCs 2, 4, 7 and 9. The English were separated at PCs 4, 5, 6, 8 and 10. The Welsh were separated at PCs 7, 9 and 10. The Finnish were separated from the other geographical locations at PCs 3 and 8 (Figure 21). As can be seen in Figures 21 and 22, many plots require analysis to draw any conclusions from the data. The facial differences are summarised in Table 6. The plots used to form this summary can be found in Appendix 2.

Table 6: Total variation explained by each PC and significance level of difference in standardised component scores (geographical location). The measurements refer to the difference between the mean shape plus/minus the respective eigenvalue/vector

PC	% Total variation	*p*-value	Facial differences for PCs that were significant following ANOVA
PC1	18.75	0.025	N/A
PC2	12.23	<0.003*	Croatian: More prominent pronasale (3mm), subnasale (2.5mm) and pogonion (2.5mm), more deep-set eyes (2.5mm)
PC3	9.49	<0.003*	Finnish: More superior nasion (3mm) and glabella (2mm), less prominent glabella
PC4	7.89	<0.003*	English: Narrower mouth (3.5mm), more superior nasion and glabella (2mm)
PC5	7.17	<0.003*	English: More superior pogonion (2mm), more inferior pronasale (1.75mm)
PC6	5.92	<0.003*	Croatian: More prominent (3.5mm) and more inferior pogonion (2mm) English: Less prominent (3.5mm) and more superior pogonion (2mm)
PC7	4.54	<0.003*	Croatian: Narrower eyes (1.5mm) Welsh: Wider eyes (1.5mm)
PC8	3.26	<0.003*	English: More prominent corners of mouth (1.5mm), more superior nasion (1.5mm) Finnish: More inset corners of mouth (1.5mm), more inferior nasion (1.5mm)
PC9	2.96	<0.003*	Croatian: More inferior pogonion (2mm) Welsh: More superior pogonion (2mm)
PC10	2.56	<0.003*	English: More inferior glabella (1mm) Welsh: More superior glabella (1mm)
PC11	2.34	0.003	English: More prominent palpebrae superius (1mm) Welsh: More inset palpebrae superius (1mm)
PC12	2.18	0.104	N/A
PC13	1.76	0.201	N/A
PC14	1.62	0.020	N/A
PC15	1.49	0.242	N/A

* Significant at 0.003 level (Bonferroni correction)

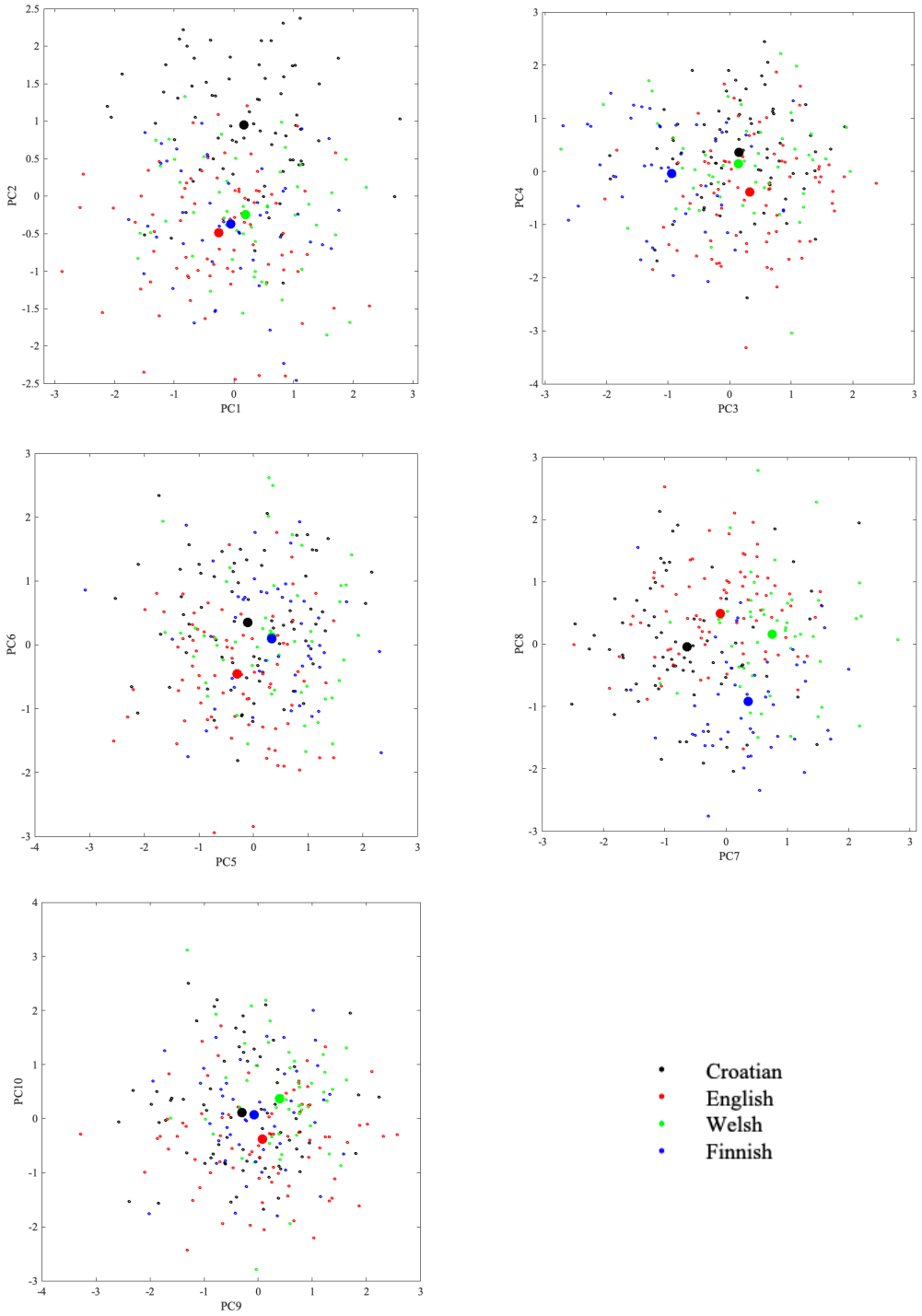


Figure 21: Scatter plots showing the group standardised component score means (labelled to highlight clustering due to geographical location) in each PC. PC1 is plotted to aid visualisation of the group separation in PC2. Interpretation of 9 PCs is required. As the influence of geographical location is contained in all 9 PCs, it challenging to ascertain the total difference between each of the groups.

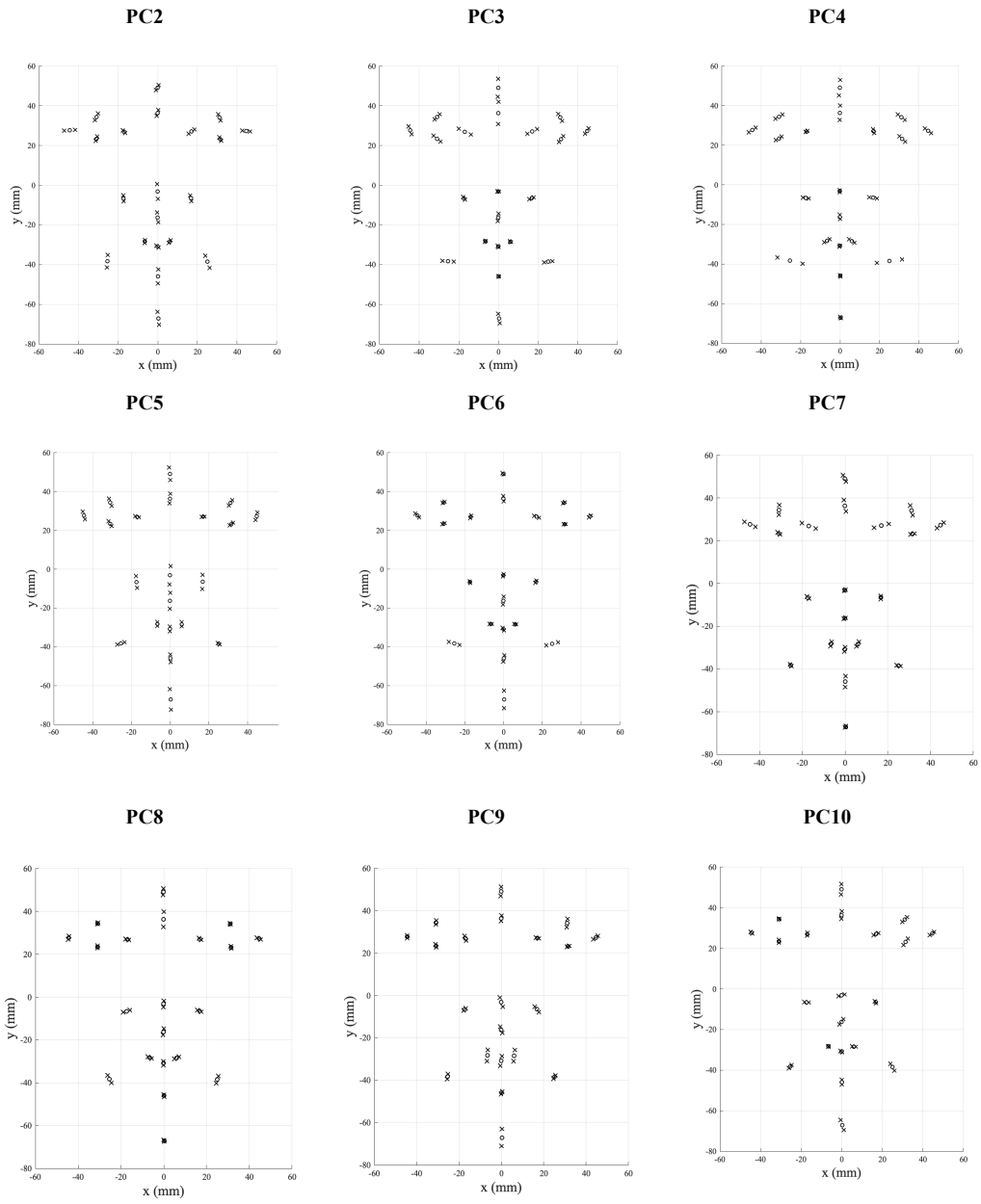


Figure 22: The number of facial plots that require interpretation when conventional PCA is used. These are presented to demonstrate the number of plots only, rather than the clinical differences which are exaggerated by a factor of four to make them easier to resolve. NB. Plots in profile also require interpretation therefore doubling the number presented here.

4.5.1.3 Discriminant Function Analysis

Along discriminate function 1 (DF1), the Croatian population was clearly separated from the other geographical locations. The English and Finnish populations were clearly separated along DF2, whilst the Welsh were distinguishable from the other populations along DF3 (Figure 23). The conventional PCs most associated with the Croatian population were PCs 2,4 and 6. PC7 and 9 were associated with the other three populations. Along DF2, conventional PC3 and 8 were associated with the English whereas PC5, 6, 7 and 10 were associated with the Finnish population. The coefficient magnitudes for DF3 were very small, with all the PCs associated with the Welsh population. Significance was found following inferential testing with Wilks Lambda (p<0.01). Interpreting the facial differences after conducting conventional PCA and DFA becomes more challenging as two different dimension reduction techniques have been used.

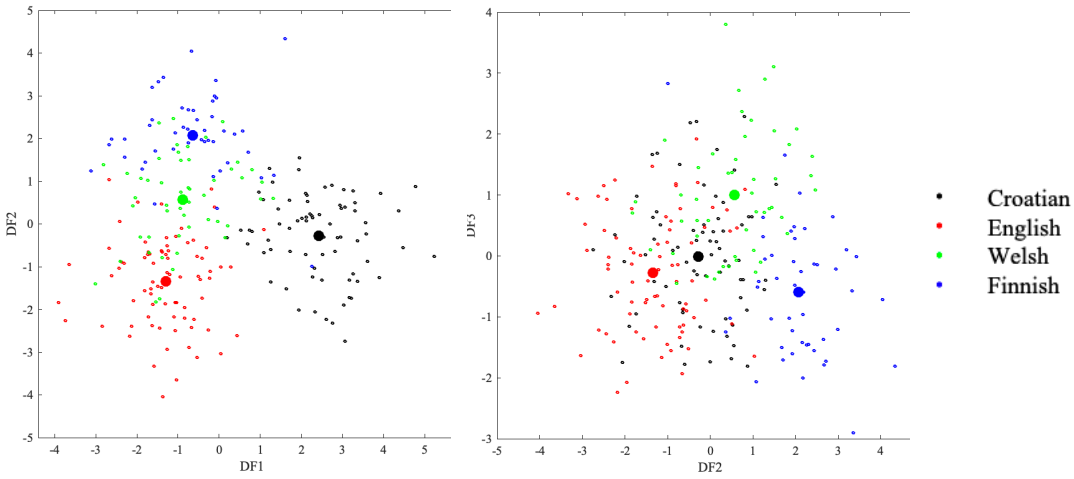


Figure 23: Scatter plots showing the group means of each of the populations in each DF after conducting conventional PCA followed by DFA. Only two plots require interpretation.

4.5.1.4 Multilevel Principal Component Analysis

As age is likely to be a confounding factor at this level, with the Croatians older than the other groups, mPCA was conducted with and without the Croatians included. When the Croatians were included in the analysis, the three retained PCs at a geographical location level explained 14.59% of the total variation. Only one PC could be retained for sex, which explained 9.98% of the total variation. Within-group variation explained the most variation. The retained 15PCs explained 61.55% of the total variation. When the Croatians were excluded from the analysis due to their older age, the two retained PCs at a geographical location level explained 11.34%, sex PC1 explained 10.93% and the 15 PCs retained at the within-group level 64.70% of the total variation. Eigenvalues via mPCA are visualised in the eigenvalue plots (Figure 24).

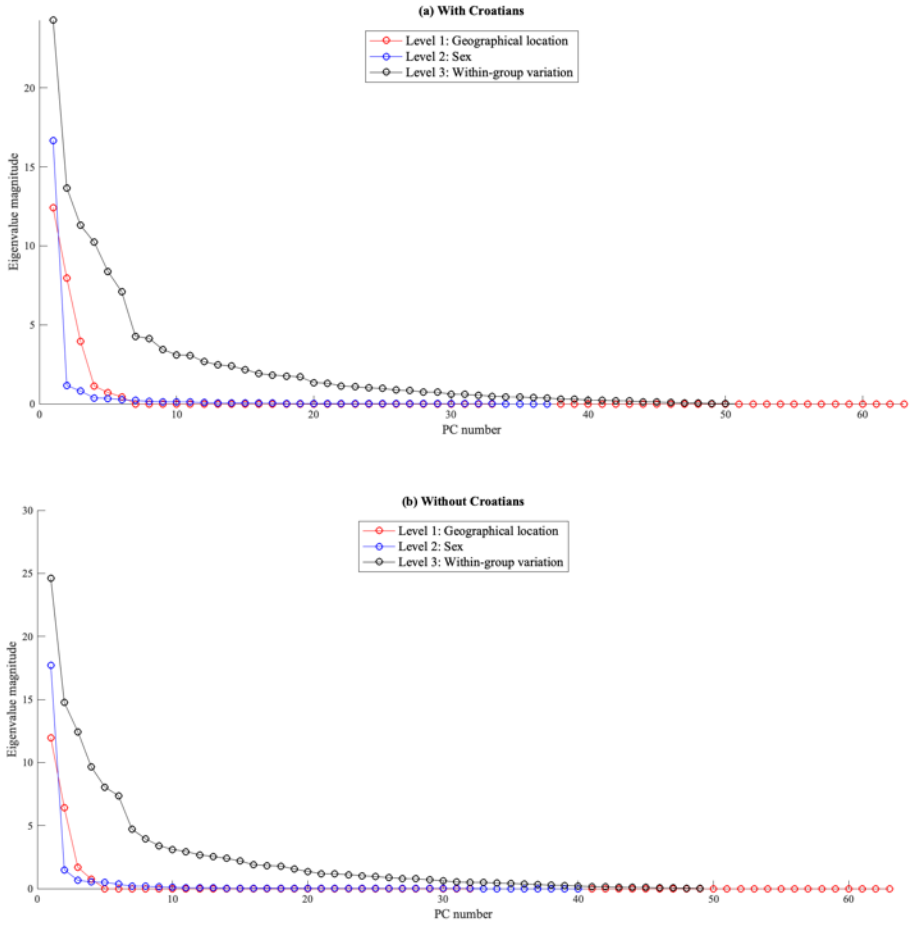


Figure 24: Eigenvalue plot for mPCA (a) with Croatians included and (b) with Croatians excluded. With the Croatians included, the importance of the geographical location eigenvalues is slightly increased.

The scatter plots present the separation of the standardised component scores for each geographical location (Figure 25). The facial differences explained by each of the PCs are summarised in Table 7. The plots that the summaries were derived from are presented in Appendix 3. As discussed previously, mPCA models were developed with and without the Croatian sample. With the Croatians excluded, subtle differences were seen in comparison to the analysis with the Croatians included (Table 7). The differences in the standardised component scores were significant (bootstrapped MANOVA applied to component scores: $p<0.001$).

With the Croatians included in the mPCA model, geographical location level PC1 explained the difference between the Croatians and the other populations (Figure 25). The largest differences were in the position of the eyes, with the Croatian population presenting with the eyes closer together (1.5-2.5mm each side). The Croatians also presented with narrower mouths, a more prominent pronasale and the corners of the eyes more inset (all approximately 2mm differences) (Table 7).

Geographical location PC2 (Croatians included) and geographical location PC1 (Croatians excluded) explained the differences between the English and Finnish samples (Figure 25). The largest differences were seen in the vertical position of nasion (2.5mm), followed by horizontal position of the corners of the mouth (2-2.5mm each side). There was little variation in profile (Table 7).

Geographical location PC3 (Croatians included) and geographical location PC2 (Croatians excluded) explained the differences between the Welsh sample and the other samples (Figure 25). Differences in the landmarks reduced considerably at this PC, with most of the variation seen in the vertical position of nasion (2mm) (Table 7).

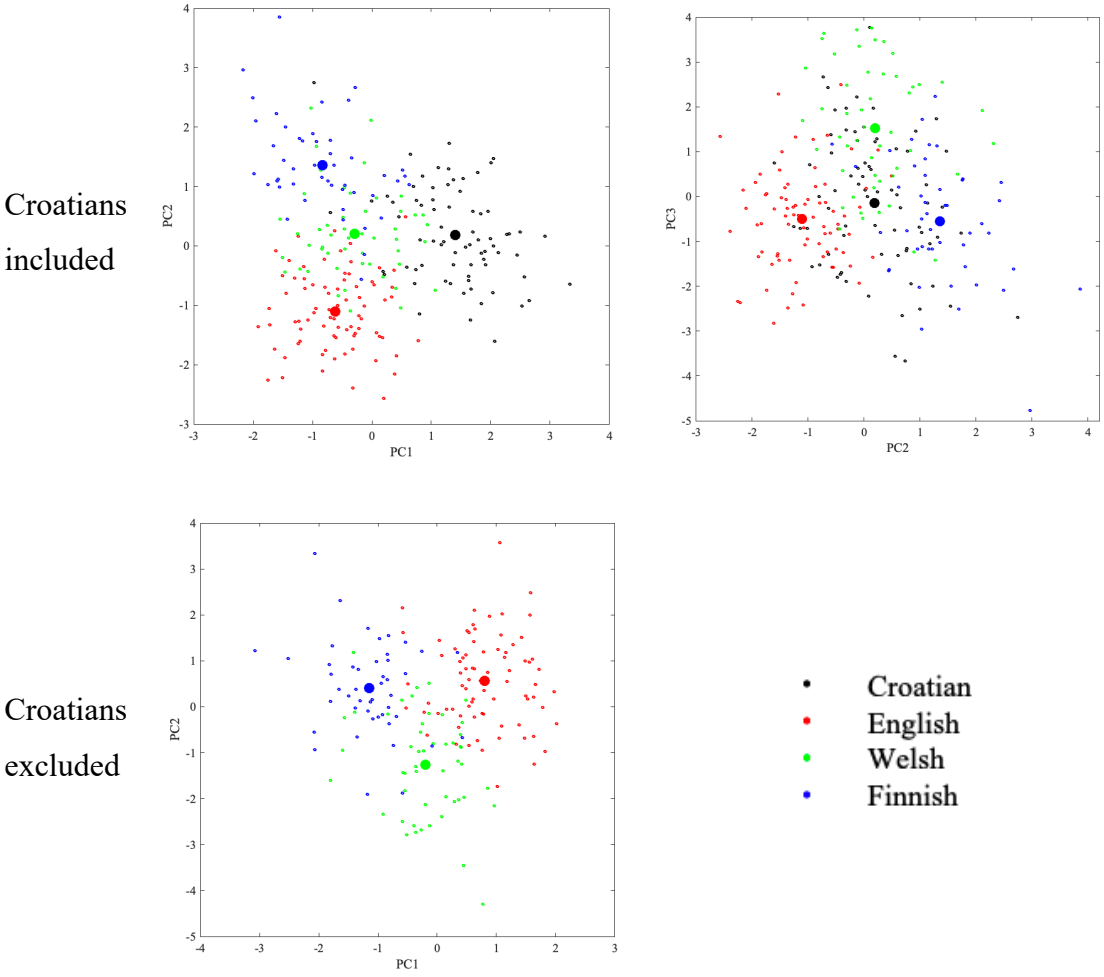


Figure 25: Scatter plots of the standardised component scores for mPCA population PC1, PC2 and PC3 (with and without the Croatians included in the model)

Table 7: Summary of the clinical differences along each geographical location PC after mPCA

Analysis	Geographical location PC	% Total variation	Summary of the differences explained by each PC where one population was isolated
With Croatians	PC1	7.44	Croatian: Eyes closer together (1.5-2.5mm each side), mouth narrower (2mm), pronasale more prominent (2mm), corner of eyes more inset (2mm) Comparison: English/Welsh/Finnish
	PC2	4.77	English: Wider mouth (2-2.5mm each side), more superior nasion (2mm) Finnish: Narrower mouth (2-2.5mm each side), more inferior nasion (2mm)
	PC3	2.38	Welsh: More superior nasion (2mm) Comparison: Croatian/English/Finnish
Without Croatians	PC1	7.39	English: Wider mouth (2-3.5mm each side), more superior nasion (2.5mm), wider, closer together and more prominent eyes (1.5-2mm) Finnish: Narrower mouth (2-3.5mm each side), more inferior nasion (2mm), narrower, further apart and less prominent eyes (1.5-2mm)
	PC2	3.95	Welsh: More superior nasion (2.5mm), more superior glabella (1.5mm), narrower mouth (1.5mm each side), more prominent pogonion (1mm)

4.5.2 Overall differences due to geographical location

In all the analyses the Croatians were most distinguishable. The findings suggest that the position and width of the eyes, width of the mouth, and nose length all differed the most between the geographical locations. Table 8 presents a summary of the differences explained by each technique.

Given that 9 PCs were explored using conventional PCA it is unsurprising there are more descriptions using this technique, however these descriptions sometimes contradict the findings of the other analyses. Most notably, the finding that the English population has a narrower mouth at PC4. Not only does this contradict the finding of mPCA and the raw landmarks, but also the finding of conventional PC1 which showed the English population having wider mouths. This finding was not displayed here as PC1 did not reach significance at a 0.003 level; however, the associated plots are available in Appendix 3.

For this reason, the results of conventional PCA are interpreted with caution. The features that are most distinguishable in each population using the raw landmark and mPCA results are presented in Figure 26.

Table 8: The largest facial differences due to geographical location as explained by each of the analyses

	Raw landmarks	Conventional PCA	mPCA
Glabella	Subtle differences	Croatian: More inferior glabella English: Contradiction Welsh: More superior glabella Finnish: More superior and less prominent glabella	Subtle differences
Eyes	Croatian: Eyes wider, closer together and more inset	Croatian: More inset eyes, narrower eyes English: More prominent palpebrae superius Welsh: More inset palpebrae superius, wider eyes	Croatian: Eyes wider, closer together, corner of eyes more inset
Nose	Croatian: Pronasale more prominent Finnish: More inferior nasion	Croatian: More inferior nasion, more prominent pronasale, more prominent subnasale English: More superior nasion, more inferior pronasale Finnish: More superior nasion, more inferior nasion	Croatian: Pronasale more prominent English: More superior nasion Welsh: More superior nasion Finnish: More inferior nasion
Mouth	English: Wider	Croatian: Wider mouth English: More prominent corners of mouth, narrower mouth Finnish: More inset corners of mouth	Croatian: Narrower English: Wider Finnish: Narrower
Chin	Subtle differences	Croatian: More prominent and more inferior pogonion English: Less prominent and more superior pogonion Welsh: More superior pogonion	Subtle differences

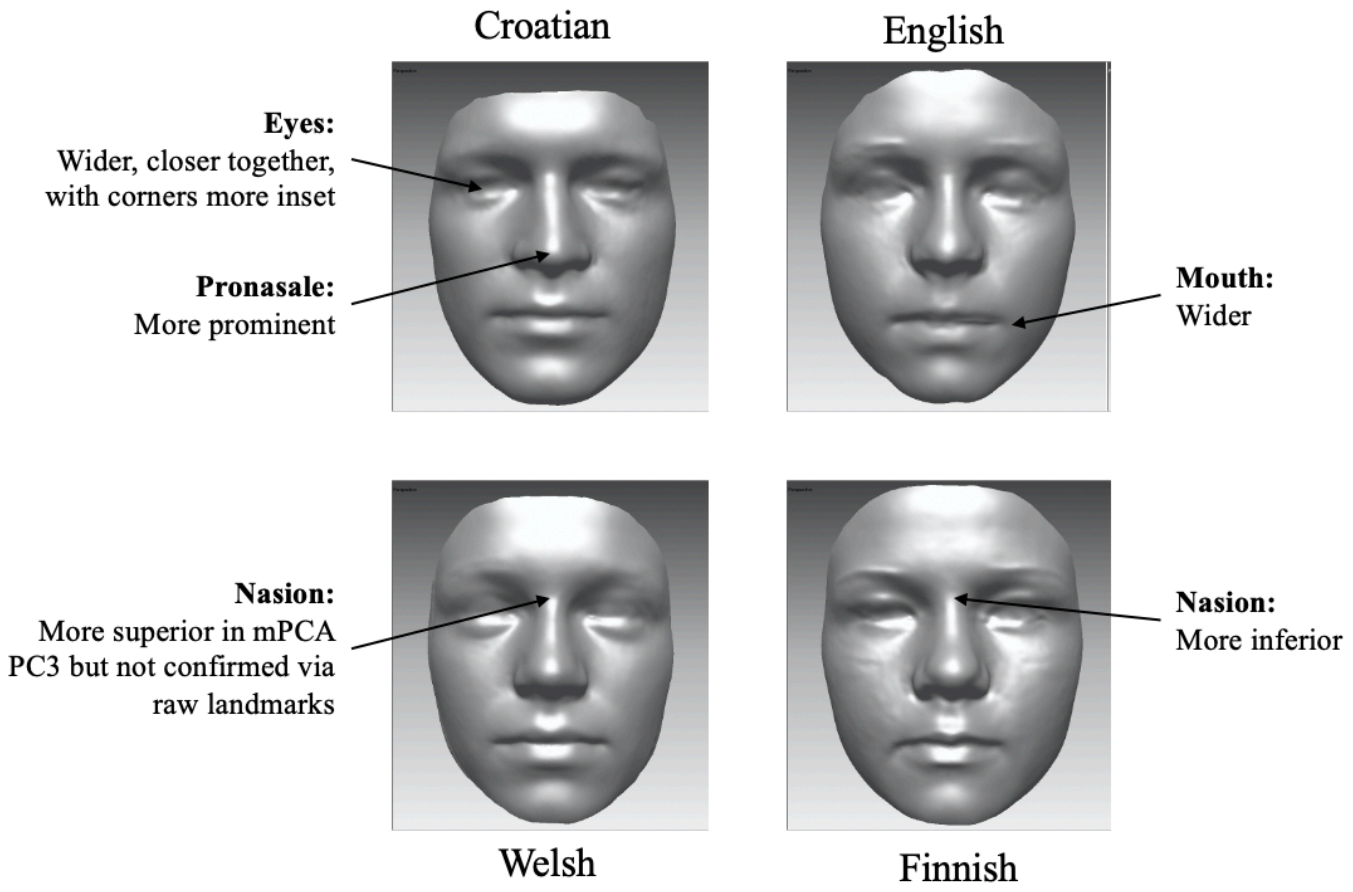


Figure 26: Overview of largest differences in facial shape due to geographical location across all analyses

4.5.3 Influence of sex

4.5.3.1 Landmarks only

The differences in the facial shapes of the male and female groups when assessing the landmarks are visualised in Figure 27. The mean of each landmark and the associated standard deviations are summarised in Appendix 2. In the transverse axis (*x*-axis), there are subtle differences between the sexes at the orbital landmarks. Females have slightly wider eyes (approximately 1mm) recorded by exocanthion position.

In the vertical axis (*y*-axis), glabella is more superiorly positioned in the female subjects. The differences are in the region of 1mm in the Croatian and English populations, and slightly more pronounced at 2mm in the Welsh and Finnish populations. The pogonion is also slightly more superiorly positioned in females and therefore the lower facial height may be slightly reduced compared to males. This appears to be isolated to the English and Croatian populations, in the region of 2mm.

The differences in the anteroposterior direction (*z*-axis) are evident in glabella, nasion, infra-orbital landmarks and pronasale. The differences in glabella are in the region of 1.5mm and are consistent across the populations with glabella more prominent in males. Sexual dimorphism in the prominence of nasion is most pronounced in the English population (1.5mm) and least pronounced in the Croatian populations (0.7mm) with nasion more prominent in the male populations. Differences in the prominence of the infraorbital landmarks are again more pronounced in the English populations (2mm). However, these differences are consistent across the populations with the infraorbital landmarks more pronounced in the female subjects. Pronasale is more prominent in the male subjects by 1mm compared to the females in all populations.

When assessing the overall mean, there appears to be a clear difference between the male and female subjects. However, as was seen with the geographical location analysis, the difference is of the magnitude of $x10^{-6}$mm.

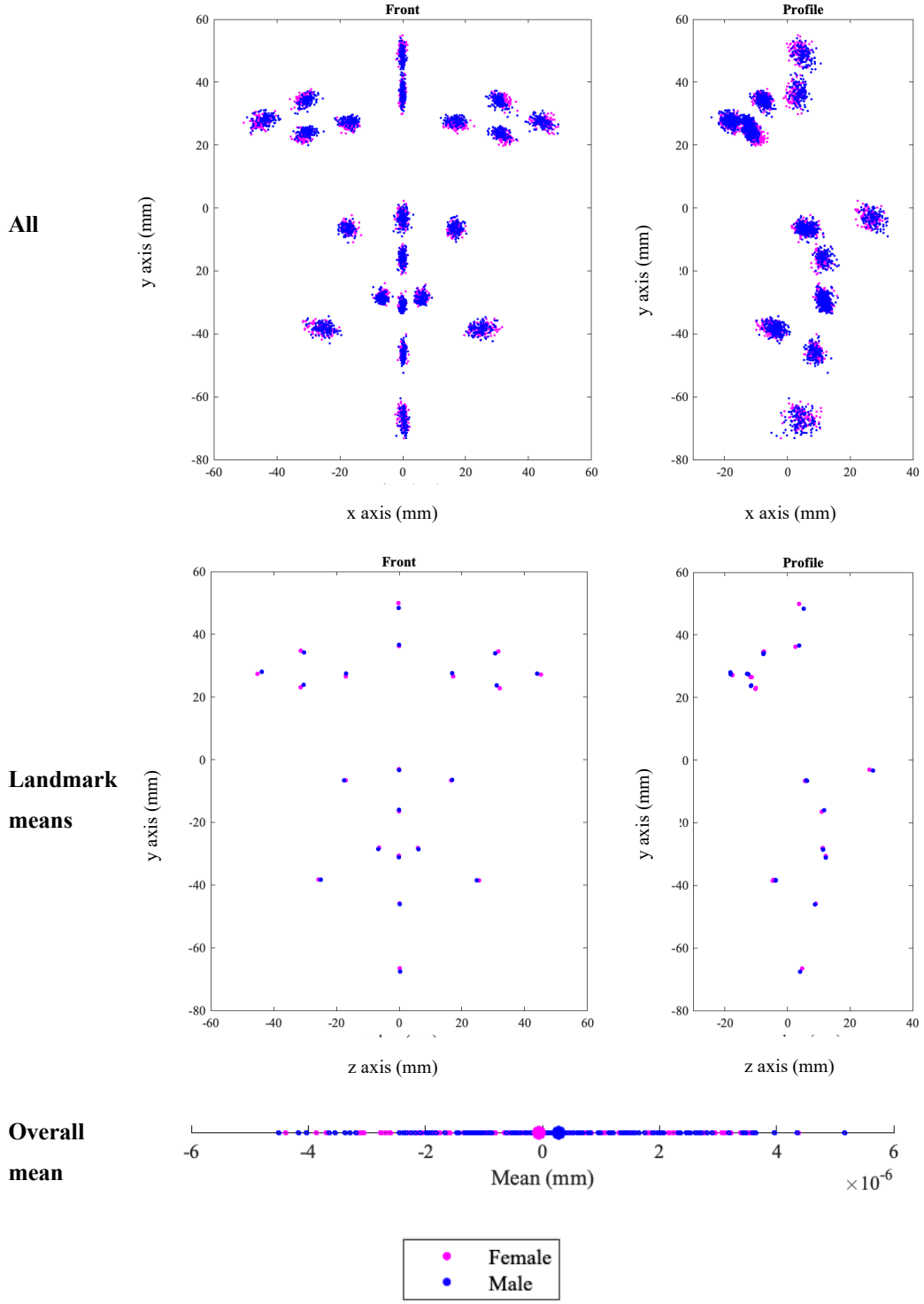


Figure 27: The 21 landmarks grouped by sex. Both between-group and within-group variation can be visualised. The main differences appear to be in the z-axis

4.5.3.2 Conventional Principal Component Analysis

The MANOVA model was found to be significant at a 0.05 level. Of the 15 PCs assessed PC1, PC3, PC5 and PC7-8 were found to be significant at a 0.003 level with regards to sex using univariate ANOVAs (Table 9). Bonferroni correction was used on the significance level to account for multiple testing. Clustering is evident in all the significant PCs (Figure 28). A summary of the facial differences explained by all the significant PCs is presented in Table 9. The plots and figures used to generate this summary can be found in Appendix 3.

Table 9: Total variation explained by each PC and significance level of difference in standardised component scores (sex)

PC	% Total variation	*p*-value (univariate ANOVAs)	Facial differences for PCs that were significant following ANOVA (females v males)
PC1	18.75	<0.003*	Wider eyes (3mm), wider mouths (2.5mm), less prominent pronasale (2mm)
PC2	12.23	0.731	N/A
PC3	9.49	<0.003*	More inferior nasion (3mm), more inferior (2mm) and less prominent glabella (2.5mm), less prominent pogonion (2mm)
PC4	7.89	0.014	N/A
PC5	7.17	<0.003*	More inferior pogonion (3mm), more superior pronasale (2mm)
PC6	5.92	0.074	N/A
PC7	4.54	<0.003*	Eyes wider apart (1.5mm)
PC8	3.26	<0.003*	More inset corners of mouth (1.5mm), more inferior nasion (1.5mm)
PC9	2.96	0.188	N/A
PC10	2.56	0.739	N/A
PC11	2.34	0.192	N/A
PC12	2.18	0.302	N/A
PC13	1.76	0828	N/A
PC14	1.62	0.531	N/A
PC15	1.49	0.797	N/A

* Significant at 0.003 level (Bonferroni correction)

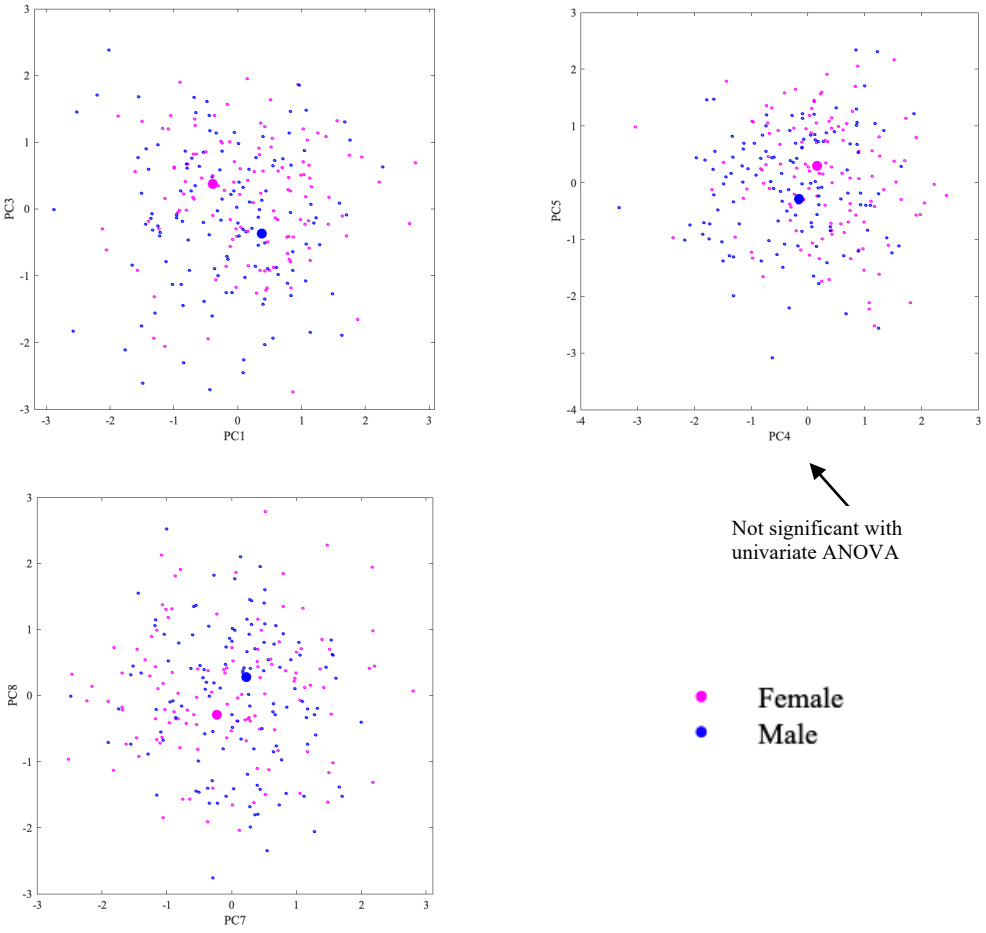


Figure 28: Scatter plots of the standardised component scores for conventional PCA PC1, PC3, PC4, PC5, PC7 and PC8 (labelled for clustering due to sex). PC4 is plotted only to help visualisation of PC5.

4.5.3.3 Discriminant Function Analysis

Only one DF is possible when assessing sex as there are two groups. Clear clustering was evident (Figure 29). As was discussed with regards to the analysis on geographical location, it is more difficult to ascertain the clinical differences in the facial shapes as two different dimension reduction techniques have been used. The differences were significant following inferential testing with Wilks Lambda (p<0.01).

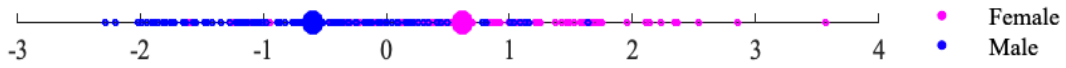


Figure 29: Visualisation of DF1 following conventional PCA

4.5.3.4 Multilevel Principal Component Analysis

With the Croatians included, sex explained 9.96% of the total variation. With the Croatians excluded, sex explained 10.92% of the total variation. At the sex level, only one PC is possible due to the rank of the matrices (as there are only two groups, only one PC is possible). The average component scores of the females and males were clearly separated in both analyses (Figure 30), demonstrating that there are clear differences between the groups. The differences in the standardised component scores were significantly different (bootstrapped ANOVA applied to component scores: p<0.001). These differences appear to be mostly in the orbital landmarks and vertical glabella position in a frontal view, with some more subtle differences in the vertical position of pogonion and horizontal position of the corners of the mouth. In profile, the variation is mostly present in glabella, nasion, pronasale and pogonion prominence. There also appears to be differences in orbital prominence, but this is harder to resolve (Figure 31).

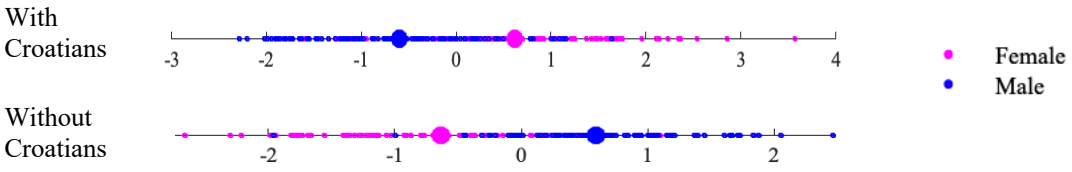


Figure 30: Scatter plots of the standardised component scores for mPCA sex PC1 (with and without the Croatians included in the model)

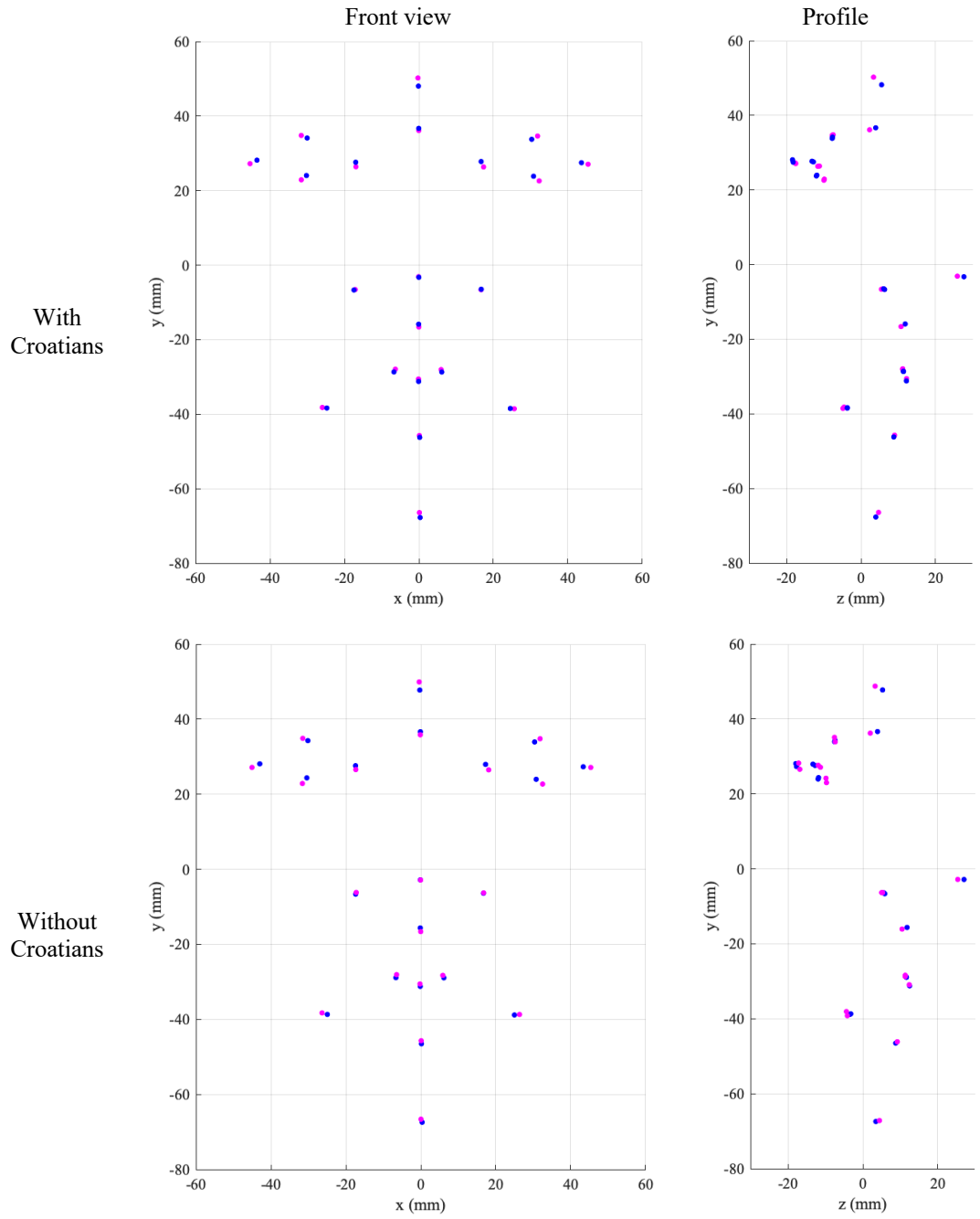


Figure 31: Visualisation of the facial differences due to sex (mPC1)

4.5.3.5 Overall differences due to sex

The differences in each of the facial features as determined by visualising the raw landmarks, using conventional PCA and mPCA are summarised in the Table 10 and Figure 32. There is a consensus between each of the analyses with the exception of the vertical position of glabella. The differences were more difficult to ascertain with conventional PCA due to the number of PCs to interpret. The differences in pogonion prominence were different in two of the PCs. The clinical differences explained by the raw landmarks and using mPCA are in agreement. This provides validity to the mPCA results.

Table 10: The facial differences due to sex as explained by each of the analyses

	Raw landmarks	Conventional PCA	mPCA
Glabella	Females more superior and less prominent	Females more inferior and less prominent	Females more superior and less prominent
Eyes	Females larger and less prominent	Females wider eyes and eyes wider apart	Females larger and less prominent
Nose	Females less prominent pronasale and nasion	Females less prominent pronasale and more inferior nasion	Females less prominent pronasale and nasion
Mouth	Females mouth wider	Females wider mouth and more inset corners of mouth	Females mouth wider
Chin	Females more superior	Females more superior Conflicting results with regards to prominence	Females more superior

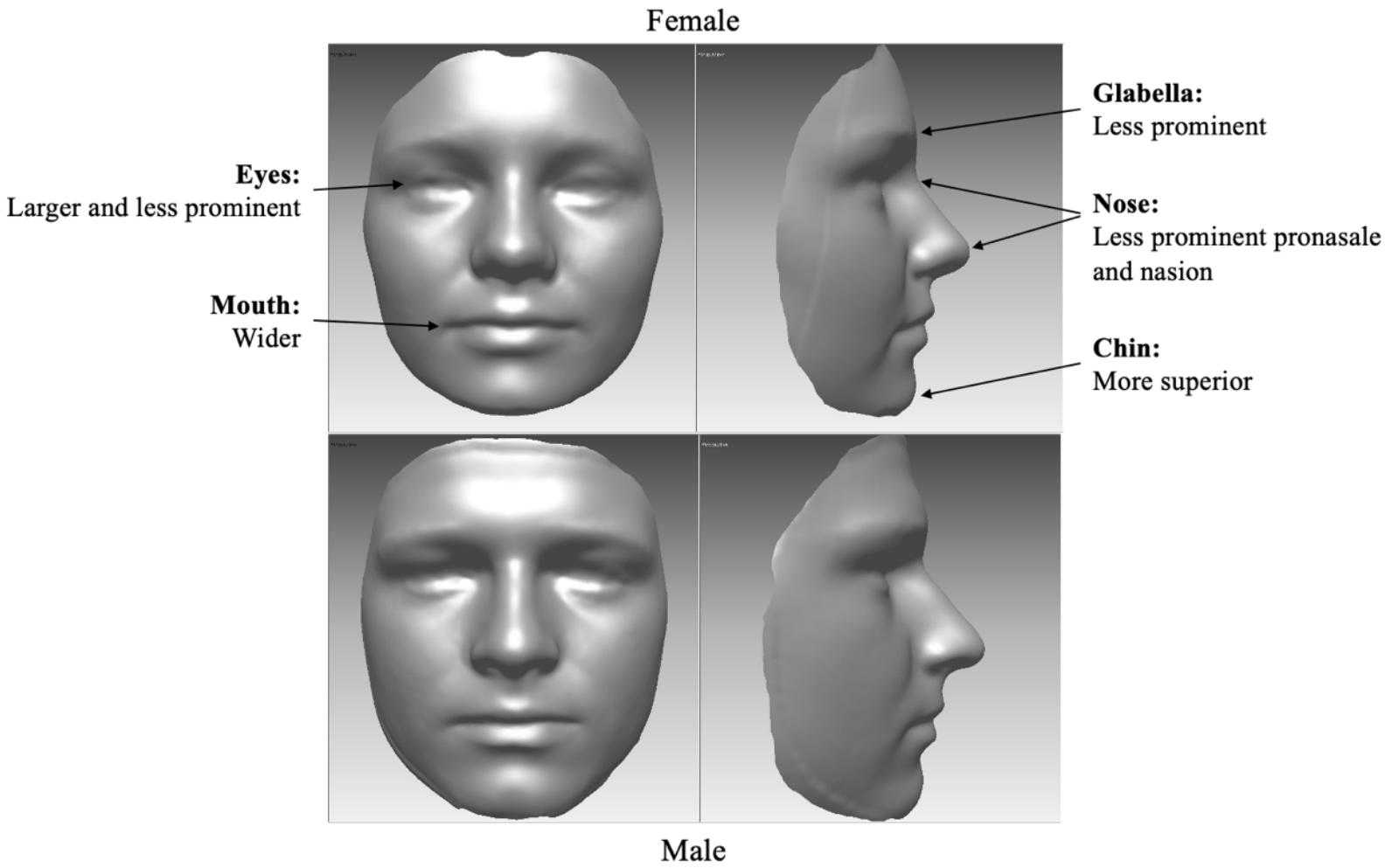


Figure 32: Overview of differences in facial shape due to sex that are confirmed across all analyses

4.5.4 Within group variation

One of the strengths of mPCA is ability to assess within-group variation (i.e., all effects apart from geographical location and sex) at a separate level of the model. When the Croatians were included, within-group variation explained 66.63%. When the Croatians were excluded, within-group variation explained 69.89% of the total variation. As this level concerns within-group variation, there is expected to be little between-group variation. It therefore follows that the group centroids are centred at the origin (p>0.05 with bootstrapped MANOVA). The small amount of variation seen may be due to slightly different group sample sizes (Figure 33).

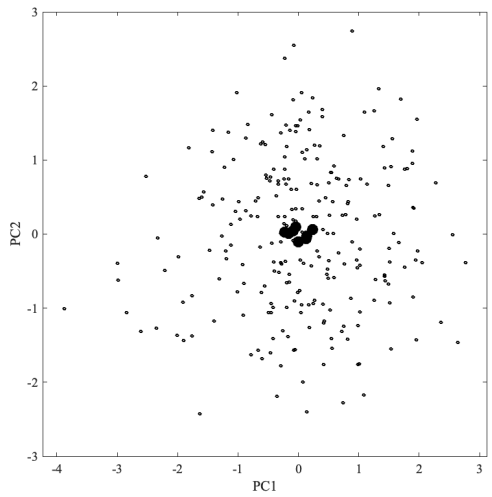


Figure 33: Standardised component scores for PC1 and PC2 (within-group variation level). Croatians included.

The facial differences explained by each of the within-group variation level PCs show very similar landmark variations to those seen in conventional PCA and highlight that most of the variation in conventional PCA is due to within group variation, rather than between group variation. All the landmarks show variation at PC1, with the main differences in the orbits (particularly in the x-axis, followed by the vertical position). The vertical position of the corners of the mouth are also clear. There was very little difference in the profile. At PC2, the vertical position of the nose, mouth and pogonion show the most variation (Figure 34). There are 22 more PCs that could be analysed at this level. However, given that the between group variation is of most interest here, the other PCs have not been visualised.

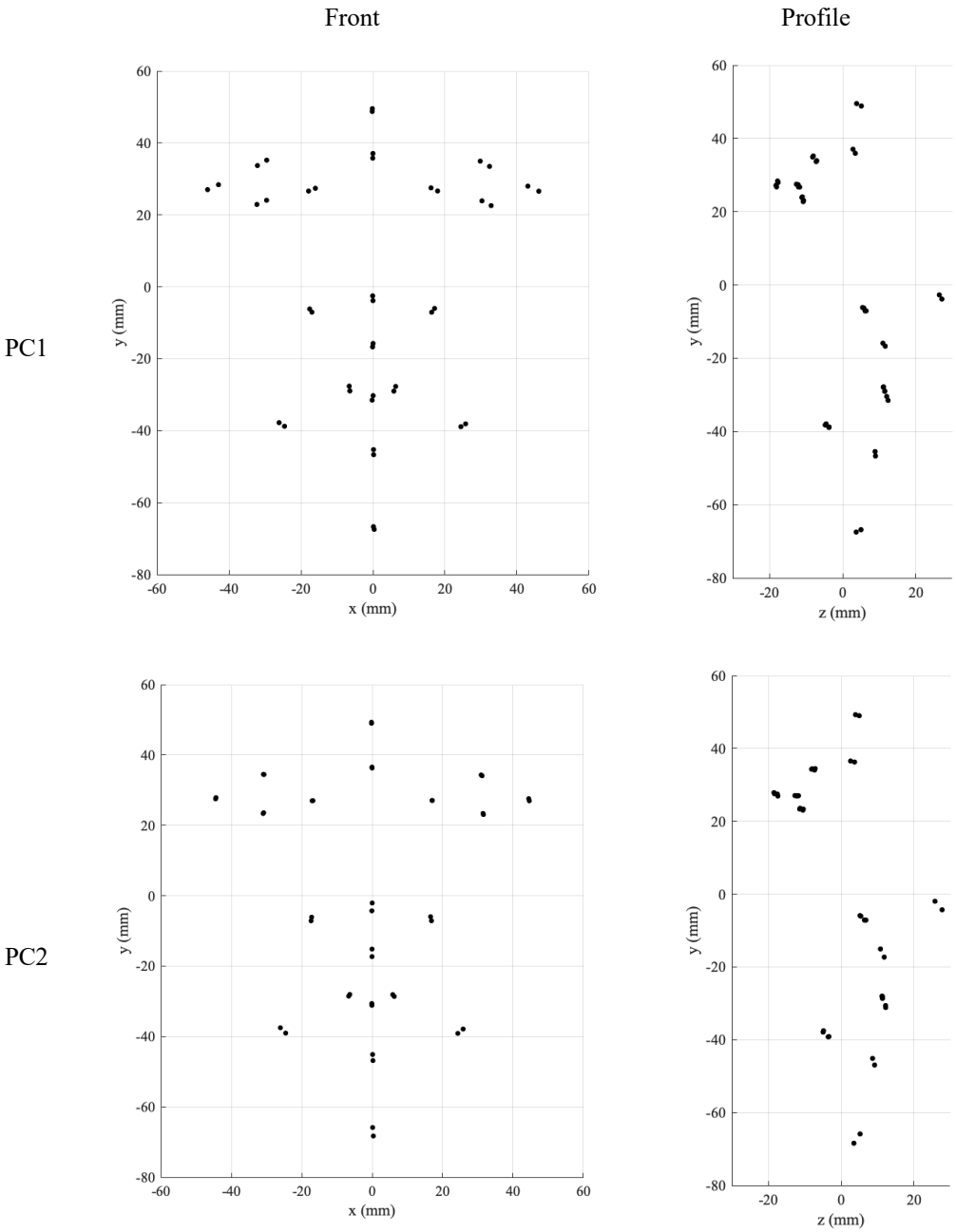


Figure 34: Interpretation of PC1 and PC2 at the within-group variation level. Mean face +/- square root of the respective eigenvalue/eigenvector.

4.6 DISCUSSION

The facial differences explained by geographical location, sex and within-group variation are discussed in General Discussion (7.1 Facial differences, p. 165). The discussion below focuses on the primary objective of this study: to assess the feasibility of mPCA prior to the investigation of more variables in the ALSPAC cohort.

The use of the landmarks alone was the easiest to assess with regards to the real clinical differences between the facial shapes of the groups as minimal mathematical techniques have been used. The plots show both between and within-group variation. However, these become more difficult to interpret as the number of groups increases. As caution is required for MANOVA, this was not implemented here.

Conventional PCA does not appear particularly useful from a visualisation perspective as the number of plots increases with the use of this technique. As the influence of geographic location, sex and within-group variation is mixed in each of the plots, it is difficult to attribute the differences visualised at each PC to one variable in isolation. Furthermore, the percentage variation explained by each of the PCs is not particularly useful as it cannot be attributed to one variable in isolation. Conventional PCA does however facilitate the use of MANOVA to determine inference.

Using conventional PCA also allows DFA to be used as the dimension reduction reduces sample sizes to within those needed to meet the assumptions of DFA. DFA increases the separation of the group means due to its supervised nature and makes it clear which groups are most distinguishable. Inference is subsequently possible via Wilks Lambda. It is however more challenging to interpret the facial differences as two mathematical techniques have been used and would require multiple steps to be resolved. The percentage of the total variance explained by each DF provides little meaningful information in the context of facial shape.

mPCA shows a very similar pattern of group mean separation to DFA, which provides validity. By using the maximum variance, the group means are separated well, again allowing easy visualisation of the groups that are most distinguishable. The

percentages highlighting the variation explained by each variable is much more meaningful than that possible with conventional PCA and DFA. However, as the maximum variation in the data set is fundamental to the technique, mPCA may be susceptible to overfitting, meaning the results may not be generalisable to other samples from the population. It is hoped that the linear nature of mPCA reduces this risk. A further disadvantage of mPCA is that the number of retained PCs is limited to the number of groups minus one. This may limit the information available for analysis, although this does also make the number of plots for visualisation more manageable.

4.7 SUMMARY

mPCA appears to be a useful tool for investigating the influence of categorical variables on facial shape research. The statistical tests used are summarised in Table 11. As mPCA can maximise the differences between groups and allow these to be visualised clearly using scatter plots, mPCA could be a useful tool for investigating variables that may have a small influence of facial shape. Furthermore, the possibility of attributing a percentage importance to each variable provides useful information for comparing the influence of different variables. Given that most of the variation was unexplained by geographical location or sex, further work is required to ascertain which other variables influence facial shape. It is for these reasons that mPCA is used in the next study to assess the influence of sex, height, BMI, age (14-16 years old), pubertal stage, metabolic factors, breathing disorders, maternal smoking during pregnancy and maternal alcohol consumption during pregnancy on the facial shape of English adolescents.

Table 11: Summary of results for study one

Test	*P*-value		Threshold for significance	Null hypothesis
	Geographical location	Sex		
Landmarks	N/A			
Conventional PCA and MANOVA	<0.001	<0.001	0.025	Reject
DFA and Wilks Lambda	<0.001	<0.01	0.025	Reject
mPCA and MANOVA	<0.001	<0.001	0.025	Reject

5 STUDY 2: USING MULTILEVEL PRINCIPAL COMPONENT ANALYSIS TO INVESTIGATE THE INFLUENCE OF MULTIPLE CATEGORICAL AND CONTINUOUS VARIABLES ON THE FACIAL SHAPE OF ENGLISH ADOLESCENTS

5.1 INTRODUCTION

In this study, mPCA will be used to explore the influence of multiple categorical and continuous variables on facial shape. Namely, sex, height, BMI, age, pubertal stage, metabolic factors, atopy, breathing disorders, maternal smoking during pregnancy and maternal alcohol consumption during pregnancy. Many of these variables have been suggested previously as having the possibility of influencing facial shape, but the methodology used in these previous studies may not have maximised the differences explained by each of the variables. It is therefore possible that the influence of these variables has previously been understated and subtle differences missed. As suggested above, mPCA is a useful tool for maximising and visualising differences between groups. It also provides a method whereby variables can be compared with regards to their relative importance in influencing facial shape. For these reasons, mPCA will be utilised here. This also provides an opportunity to assess the ability of mPCA to investigate continuous variables as well as categorical variables.

5.2 AIMS

- To explore the suitability of mPCA as an analytical tool for both categorical and continuous variables.
- To determine the influence of sex, height, BMI, age, pubertal stage, metabolic factors, breathing disorders, maternal smoking during pregnancy and maternal alcohol consumption during pregnancy on the facial shape of English adolescents using mPCA.

5.3 NULL HYPTHESES

- Sex is not associated with the facial shape of English 14 to 16 year olds.
- Height is not associated with the facial shape of English 14 to 16 year olds.
- BMI is not associated with the facial shape of English 14 to 16 year olds.
- Age (14-16 years old) is not associated with the facial shape of English 14 to 16 year olds.
- Pubertal stage (Tanner Stage 3-5 due to sample sizes) is not associated with the facial shape of English 14 to 16 year olds.
- Metabolic factors (fasting insulin, fasting glucose, cholesterol, triglycerides, HDL, LDL and VLDL) are not associated with the facial shape of English 14 to 16 year olds.
- Atopy is not associated with the facial shape of English 14 to 16 year olds.
- Breathing disorders (asthma from 0-3.5 years old, asthma at 7.5 years old and hay fever) are not associated with the facial shape of English 14 to 16 year olds.
- Maternal smoking before pregnancy and/or during the 1st or 2nd trimesters is not associated with the facial shape of English to 14 to 16 year olds.
- Maternal alcohol consumption before pregnancy and/or during the 1st or 2nd trimesters is not associated with the facial shape of English 14 to 16 year olds.

5.4 METHODOLOGY OVERVIEW

An overview of the methodology used in this study is provided below.

5.4.1 3D facial scan acquisition, processing and landmarking

The 3D facial scans were acquired and processed as documented previously. GPA was conducted on the 3D facial scans. Scaling of the images removed the influence of size, thus isolating the influence of shape (3.4. Facial scan processing (p. 53). Initial analyses of the eigenvalue magnitudes utilised 21 manually placed landmarks (3.5 Landmarking, p. 52). The landmarks were placed by one researcher who underwent a calibration exercise controlled by an external researcher. More detailed analyses involved 1000 quasi-landmarks placed automatically using a publicly available MATLAB algorithm (3.5 Landmarking, p. 52).

5.4.2 Variables

Twenty-three variables were assessed in this chapter. Information on the collection of these variables is detailed in 3.6 Variables (p. 56). The outcome variables were initially 21 manually placed landmarks then 1000 automated quasi-landmarks.

5.4.3 Data cleaning and exploration

Subjects were excluded if they had obvious craniofacial dysmorphology or their 3D facial scan was of insufficient quality. Twins and triplets were excluded alongside those with missing data. Subjects reporting a pubertal stage of two or three were excluded due to small sample sizes (3.7.1 Exclusions, p. 58). The 21 manually placed landmarks were normally distributed (3.7.2. Distribution of landmark data, p. 59). Multicollinearity was assumed as this was present for the landmark data in study one. Although outliers were detected in the 21 landmarks, the subjects were included following inspection of the facial shells and assessing the data set without exclusions due to missing data (3.7.3.1 Outliers in landmark data, p. 62). No outliers were detected at a 4SD level for the quasi-landmarks . Outliers in the variable data were categorised into the nearest grouping during the discretisation process (3.7.5 Discretisation, p. 71) and therefore did not pose an issue. Correlations were present between some of the 23 variables assessed (3.7.4.2 Correlation between the variables, p. 66). The implications of this are discussed later.

5.4.4 Final sample

The English population that the sample is drawn from is described in 3.2 Populations (p. 51). In total, 1411 subjects were included in the mPCA models. The final categories used in the mPCA models following discretisation are detailed in Tables 12 and 13. Sample size calculations were not possible due to the novelty of the methodology.

Table 12: Descriptive analysis, sample sizes and categories for the data that was originally continuous

	Min	Max	Mean	5% Trimmed Mean	Median	SD	Distribution	**Category**	Sample Size	%
Age (months)	174.00	202.00	184.46	184.29	184.00	2.550	Right skewed	**<183**	536	38.0
								184 - 186	660	46.8
								187 - 189	156	11.1
								190 - 192	35	2.5
								>193	24	1.7
Height (cm)	145.00	200.10	169.58	169.403	169.00	8.316	Normal	**145 – 154**	34	2.4
								155 - 164	398	28.2
								165 - 174	611	43.3
								175 - 184	310	22.0
								185+	58	4.1
BMI	14.700	39.647	21.465	21.203	20.745	3.461	Right skewed	**Underweight**	219	15.5
								Healthy	1001	70.9
								Overweight	153	10.8
								Obese	38	2.7
Cholesterol (mmol)	1.250	6.870	3.765	3.740	3.730	0.643	Normal	**<=3.00**	149	10.6
								3.01 – 3.50	349	24.7
								3.51 – 4.00	466	33.0
								4.01 – 4.50	290	20.6
								4.51 – 5.00	106	7.5
								>= 5.01 (Unhealthy[1])	51	3.6
Triglycerides (mmol)	0.130	4.600	0.823	0.740	0.740	0.347	Right skewed	**<0.70**	615	43.6
								0.70 – 0.89	358	25.4
								0.90 – 1.09	207	14.7
								1.10 – 1.29	106	7.5
								1.30 – 1.49	57	4.0
								1.50 – 1.69	30	2.1
								>=1.70 (Unhealthy[1])	38	2.7
VLDL (mmol)	0.059	2.100	0.376	0.360	0.338	0.159	Right skewed	**<=0.20**	83	5.9
								0.21 – 0.40	875	62.0
								0.41 – 0.60	339	24.0
								0.61 – 0.80	89	6.3
								>0.80	25	1.8
LDL (mmol)	0.230	5.467	2.095	2.069	2.045	0.550	Normal	**< 1.50**	177	12.5
								1.51 – 2.00	485	34.4
								2.01 – 2.50	452	32.0
								2.51 – 3.00	223	15.8
								>3.00 (Unhealthy1)	74	5.2
HDL (mmol)	0.550	2.870	1.293	1.283	1.260	0.291	Normal	**<=1.00 (**Unhealthy for males[1])	212	15.0
								1.01 – 1.20 (Unhealthy for females[1])	367	26.0
								1.21 – 1.40	370	26.2
								1.41 – 1.60	261	18.5
								1.61 – 1.80	129	9.1
								>1.80	72	5.1
Glucose (mmol)	3.500	6.800	5.200	5.198	5.200	0.344	Normal	**<4.5**	19	1.3
								4.5 – 4.9	457	32.4
								5.0 – 5.4	720	51.0
								5.5 – 6.9 (Prediabetes[2])	215	15.2
Insulin (iu/l)	1.517	62.170	10.067	9.676	9.135	5.049	Right skewed	**<5.00**	159	11.3
								5.00 – 9.99	663	47.0
								10.00 – 14.99	400	28.3
								15.00 – 19.99	132	9.4
								20.00 – 24.99	45	3.2
								>=25.00 (Unhealthy[3])	12	0.9

[1]Heart UK (2021) [2]Editor (2019) [3](Buppajarntham 2019)

Table 13: Descriptive analysis, sample sizes and categories for the data that was originally categorical

	Original Category	**Category Used in Analysis**	Sample Size	%
Sex	Male	**Male**	636	45.1
	Female	**Female**	775	54.9
Pubertal Stage (Hair)	Stage 1	Excluded (n=2)		
	Stage 2	Excluded (n=8)		
	Stage 3	**Stage 3**	59	4.2
	Stage 4	**Stage 4**	666	47.2
	Stage 5	**Stage 5**	686	48.6
Pubertal stage (Genitals)	Stage 1	Excluded (n=5)		
	Stage 2	Excluded (n=14)		
	Stage 3	**Stage 3**	147	10.4
	Stage 4	**Stage 4**	733	51.9
	Stage 5	**Stage 5**	531	37.6
Asthma (0-42 months)	Never wheezed	**Never wheezed**	1068	75.5
	Wheezed at 6 months but not 42 months	**Wheezed at 6 months but not 42 months**	171	12.1
	Never wheezed 6 months but wheezed 42 months	**Never wheezed 6 months but wheezed 42 months**	121	8.6
	Wheezed both 6 and 42 months	**Wheezed both 6 and 42 months**	55	3.9
Asthma (91 months)	No	**No**	1250	88.6
	Yes	**Yes**	161	11.4
Atopy	No	**No**	1125	79.7
	Yes	**Yes**	286	20.3
Hay fever	No	**No**	1290	91.4
	Yes	**Yes**	121	8.6
Alcohol before pregnancy	0	**No alcohol**	66	4.7
	< 1 glass/week	**< 1 glass/week**	510	36.0
	1+ glasses/week	**Weekly (1+ glasses)**	656	46.4
	1-2 glasses/day	**Daily (1+ glasses)**	168	13.0
	3-9 glasses/day	Combined with 1-2 glasses/day (n=15)		
Alcohol 1st trimester	0	**No alcohol**	625	44.2
	< 1 glass/week	**< 1 glass/week**	582	41.1
	1+ glasses/week	**Weekly (1+ glasses)**	187	13.2
	1-2 glasses/day	**Daily (1+ glasses)**	21	1.5
	3-9 glasses/day	Combined with 1-2 glasses/day (n=3)		
Alcohol 2nd trimester	0	**No alcohol**	643	45.4
	< 1 glass/week	**< 1 glass/week**	539	38.1
	1+ glasses/week	**Weekly (1+ glasses)**	215	15.2
	1-2 glasses/day	**Daily (1+ glasses)**	18	1.3
Smoking before pregnancy (cigs/day)	0	**0**	1124	79.4
	1-4	**1-4**	43	3.0
	5-9	**5-9**	52	3.7
	10-14	**10-14**	54	3.8
	15-19	**15-19**	62	4.4
	20-24	**20+**	80	5.6
	25-29	Combined into 20+ cigs/day (n=12)		
	30+	Combined into 20+ cigs/day (n=9)		
Smoking 1st trimester (cigs/day)	0	**0**	1216	85.9
	1-4	**1-4**	62	4.4
	5-9	**5-9**	47	3.3
	10-14	**10-14**	42	3.0
	15-19	**15-19**	31	2.2
	20-24	**20+**	17	1.2
	25-29	Combined into 20+ cigs/day (n=4)		
	30+	Combined into 20+ cigs/day (n=1)		
Smoking 2nd trimester (cigs/day)	0	**0**	1260	89.0
	1-4	**1-4**	42	3.0
	5-9	**5-9**	39	2.8
	10-14	**10-14**	35	2.5
	15-19	**15-19**	21	1.5
	20-24	**20+**	18	1.3
	25-29	Combined into 20+ cigs/day (n=3)		

5.4.5 Analyses

mPCA is detailed in 2.9 Analysis of 3D facial landmarks (p. 40). Separate two-level, nested mPCA models were used to assess the influence of each of the 23 variables on the facial shape of the ALSPAC cohort. These were run using in-house mPCA code in MATLAB 2017b. The variable of interest was placed at level one of the models and within-group variation was placed at level two (Figure 35). The median was used for the covariance matrices.

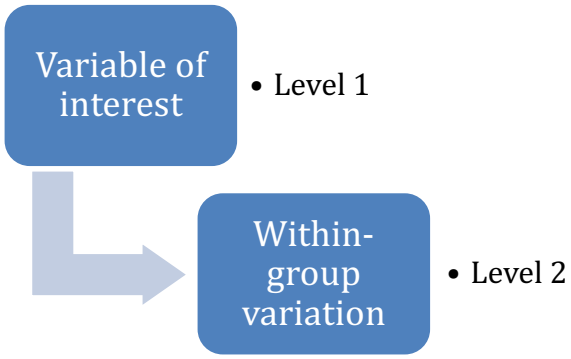


Figure 35: Two-level mPCA model structure

The importance of the each of the variables was assessed using the percentage of total variation explained by level one. This was visualised as a bar chart in a similar fashion to a Manhattan plot. Given that the 1000 landmark analyses explained more variation in general than the 21 landmark analyses, the 1000 landmark analyses were explored in further detail. The component scores were standardised by subtracting the mean and dividing by the square root of the respective eigenvalue (standard deviation). These were visualised via scatter plot in MATLAB R2017b. Inference was determined via ANOVA or MANOVA as appropriate with an additional bootstrapping step as outlined in study one (4.4.5.4 mPCA, p. 76). Following Bonferroni correction to adjust for multiple testing (23 variables), significance was set at $p<0.002$.

The facial differences explained by each of the variables were visualised using the 3D viewer class in MATLAB R2017b. Differences between the average face plus/minus the square root of the respective eigenvalue and eigenvectors were visualised.

5.5 RESULTS

5.5.1 Percentage of total variation

For most of the variables, the variation explained using 21 landmarks was less than the variation using 1000 landmarks. This is to be expected for variables that are likely to influence the deposition of adipose tissue as opposed to the main facial features (eyes, nose and mouth). To explore as much of the facial shape differences as possible, the 1000 landmark models were explored in further detail in this chapter rather than the 21 landmark models.

Of the variables explored, four variables explained more than 5% of the total variation when 1000 landmarks were used in the mPCA models (sex, height, BMI and insulin). A further seven variables (triglycerides, cholesterol, VLDL, LDL, maternal smoking during the first trimester, maternal smoking during the second trimester and maternal alcohol consumption during the first trimester) explained more than 2% of the total variation (Figure 36).

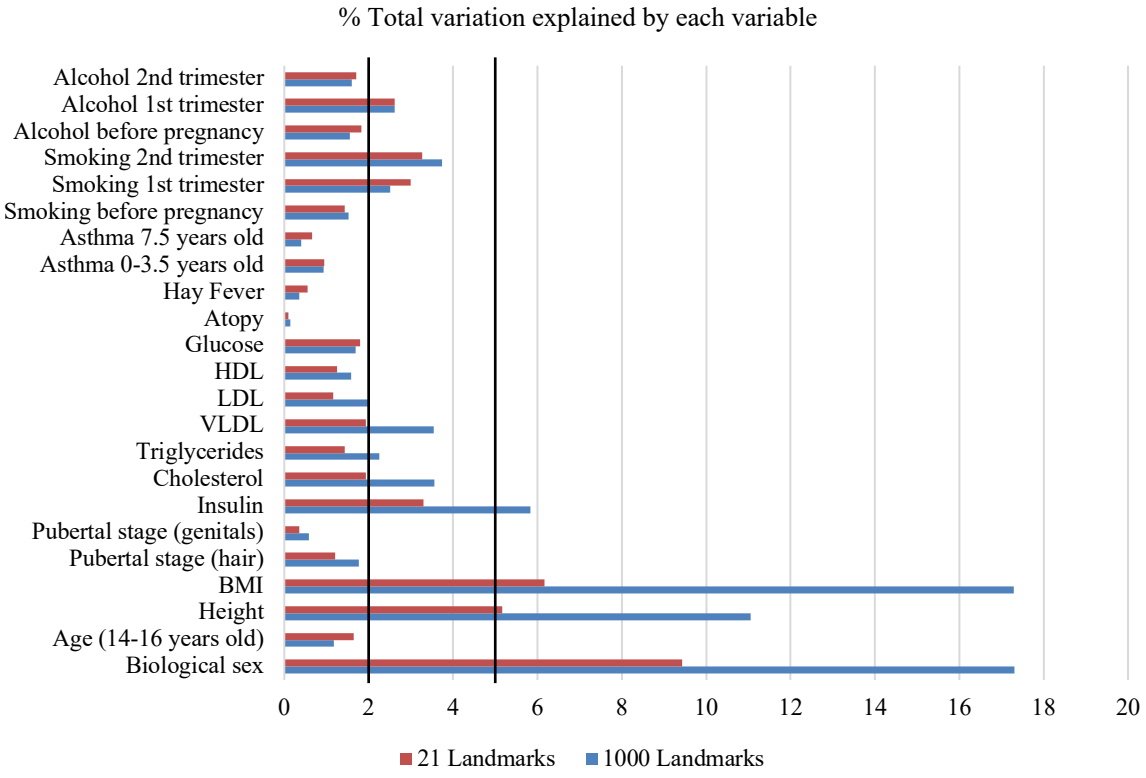


Figure 36: Percentage of the total variation explained by each variable in its own two-level analysis using both 21 landmarks and 1000 quasi-landmarks.

5.5.2 Sex

Sex explained 17.31% of the total variation in the 1000 quasi-landmark model. This was the largest of all the variables investigated. The component scores were clearly separated along PC1 with the component scores significantly different between the two groups (p<0.001).

The facial shape most representative of the female subjects had more prominent cheeks and therefore a round facial shape. The midface was relatively retrusive by comparison. The eyes were also more prominent in the female faces. By contrast, the male faces had a more prominent midface in comparison to the cheeks. This gives a narrower facial shape. The brow-ridge and nose were larger in the male faces. The shape of the chin in the male faces may represent a more prominent chin or a relatively larger facial shape, which was constrained by the scaling procedures carried out when processing of the faces. This scaling procedure limits the information available on the magnitude of the differences. It is therefore best to determine the differences relative to each other (Figure 37).

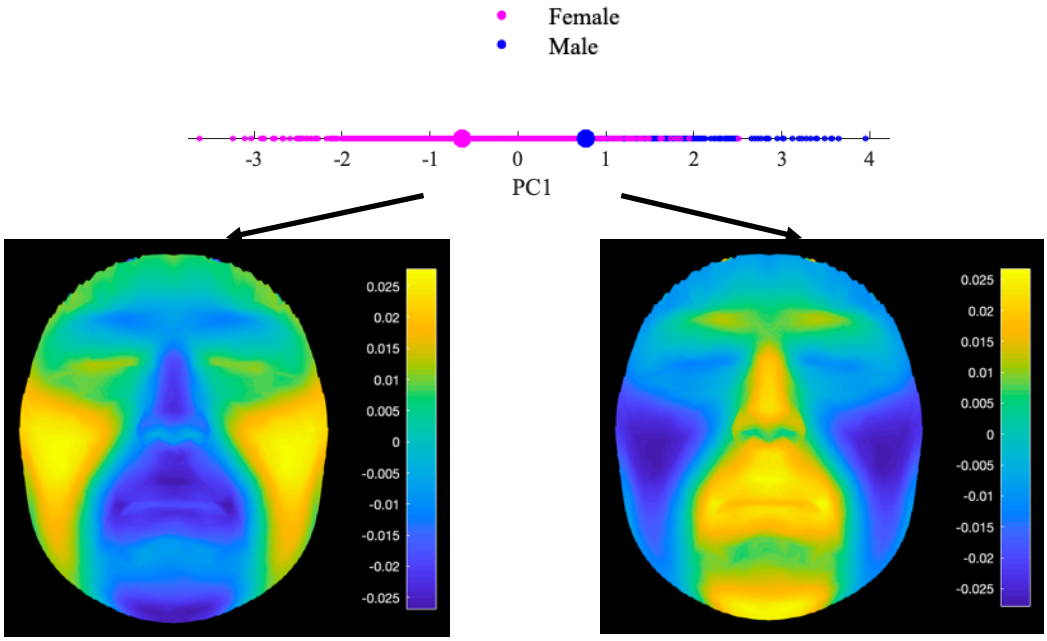


Figure 37: Visualisation of the component scores and facial differences due to sex PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.3 Age (14-16 years old)

Age explained 1.18% of the total variation in the 1000 quasi-landmark model. The separation of the component scores followed no ordered pattern, with the latter two groupings at opposite ends of the PC1 axis. At PC2, again there is no ordered separation. The component scores were not significantly different (p>0.05). Older subjects appear to have larger supra-orbital regions, noses, mouth and chins at PC1 but this is not replicated in the results of PC2. Along both PC1 and PC2, the groups with the smallest sample size are the groups that are most separated from the other groups. It is unclear whether the separation is due to true differences or whether this is due to a discrepancy in the sample sizes (Figure 38).

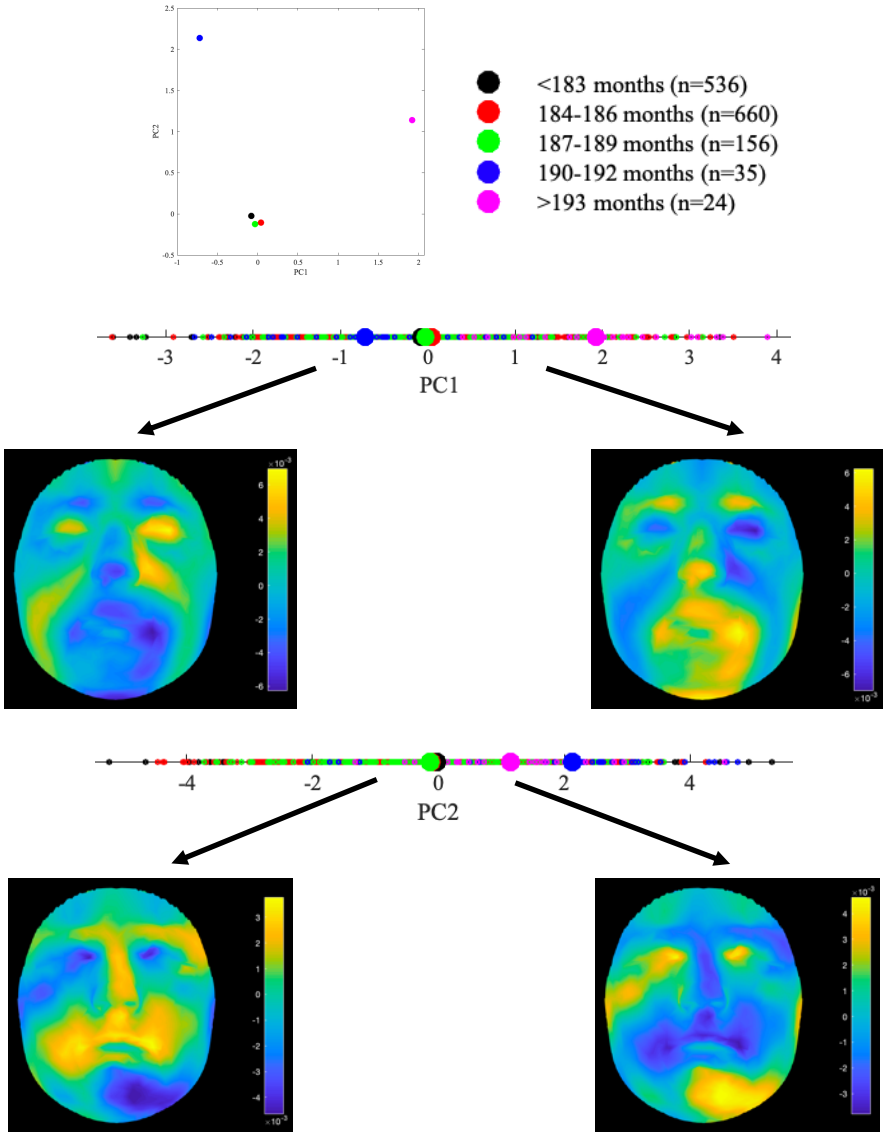


Figure 38: Visualisation of the component scores and facial differences due to age (14-16 years old) PC1 and PC2. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.4 Height

Height explained 11.05% of the total variation in the 1000 quasi-landmark model. The separation of the component scores along PC1 followed an ordered pattern with shorter subjects presenting with negative mean component scores and taller subjects with positive mean component scores. There appears to be equal separation of the groups despite differing sample sizes. The separation of the component scores was significant (p<0.001). Those of smaller stature had more prominent cheeks relative to the midface and therefore a rounder facial appearance. Those with a taller stature had a relatively narrower facial shape with larger midface including the lips, nose and chin, and to a lesser extent, brow-ridge. Further interpretation is required to ascertain if these differences are in the x-, y- or z- axis (Figure 39).

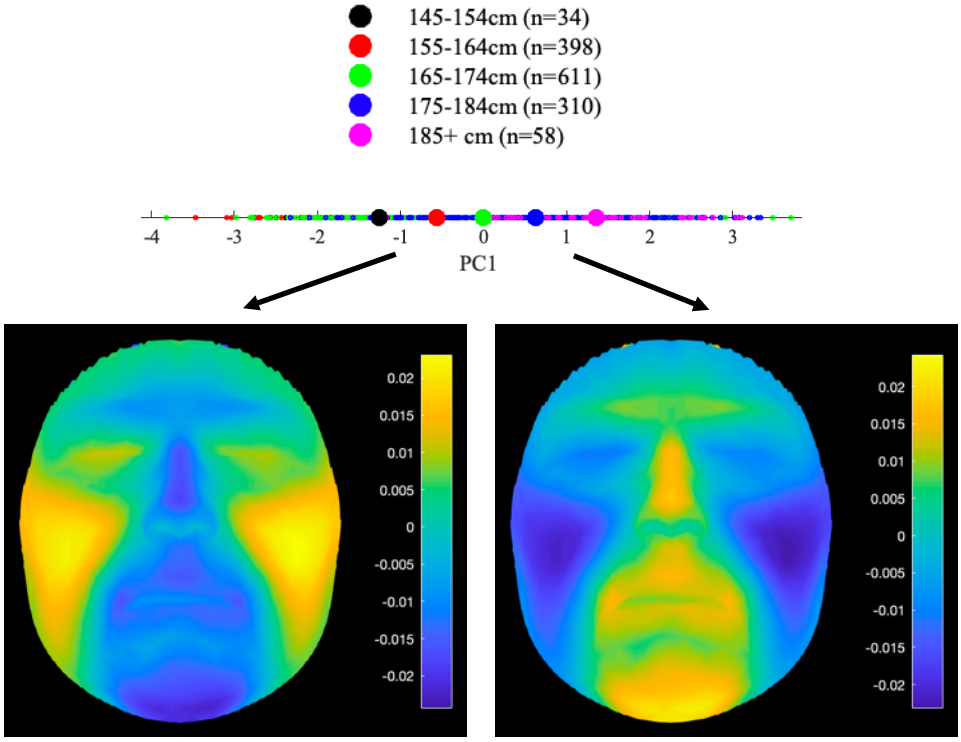


Figure 39: Visualisation of the component scores and facial differences due to height PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.5 BMI

BMI explained 17.29% of the total variation in the 1000 quasi-landmark model. The separation of the component scores along PC1 followed an ordered pattern with underweight subjects presenting with a negative mean component score, whilst overweight and obese subjects produced positive mean component scores. The separation of the component scores was significant (p<0.001). The group with the smallest sample size, obese, is separated further than the other groups. Those that were underweight had less prominent cheeks and a relatively more prominent midface. The forehead and supra-orbital region also appear to be influenced. Those that were overweight or obese presented with more prominent cheeks and a relatively retrusive midface and/or reduced mid and lower facial height. The forehead also appears to be influenced in a positive direction but to lesser extent than the cheeks. Due to the scaling procedures, it is difficult to determine whether the differences at the facial periphery were due to facial shape differences or represent differing overall facial size but constrained by the scaling procedures (Figure 40).

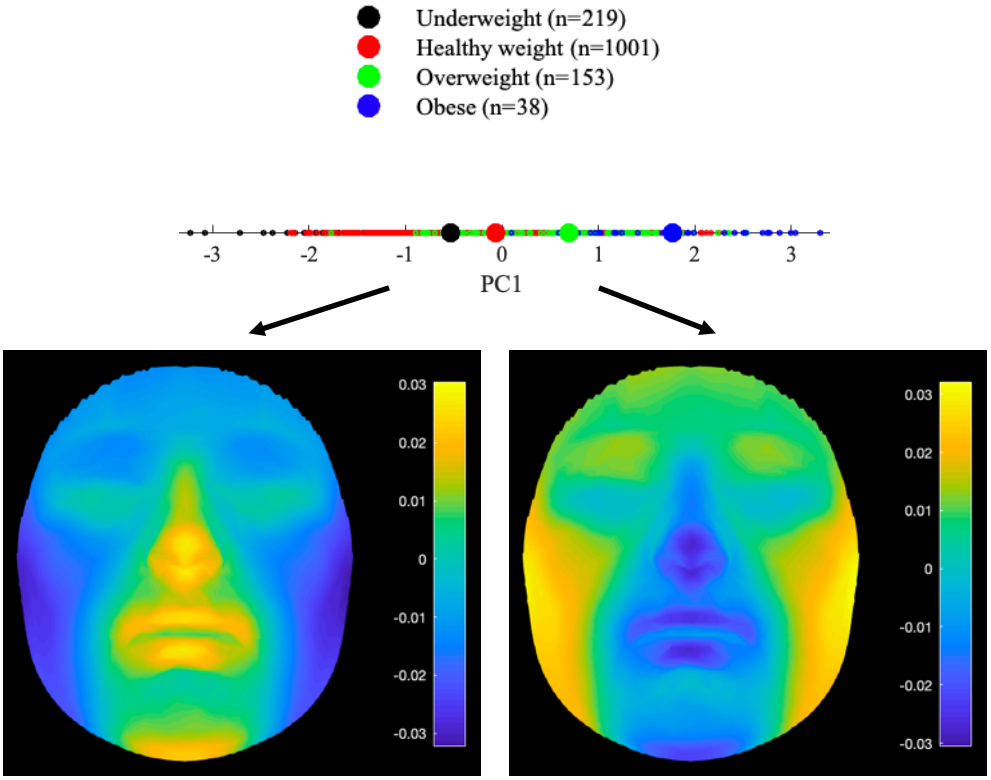


Figure 40: Visualisation of the component scores and facial differences due to BMI PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.6 Pubertal stage (pubic hair)

Pubertal stage as determined through self-declared Tanner stage with regards to pubic hair explained 1.76% of the total variation. The component scores show an ordered separation along PC1. The mean component score of subjects at stage three of puberty were negative, whilst the mean component scores of subjects at stage four and five were near the origin. Little separation was seen been the mean component scores of subjects at stage four and stage five. Component scores were significantly different (p<0.001). Stage three had the smallest sample size.

Those reporting stage three development appeared to have more prominent cheeks and upper lip, with a relatively more retrusive chin. By contrast, those reporting stage four or five development had more retrusive cheeks with a relatively larger chin. The brow-ridge may also have been slightly more prominent. Although, it is not possible to determine absolute magnitudes of the differences due to the scaling process, it should be noted that the magnitude of the differences is smaller than those seen for sex, height and BMI, which is reflected by the percentage of total variation explained (Figure 41).

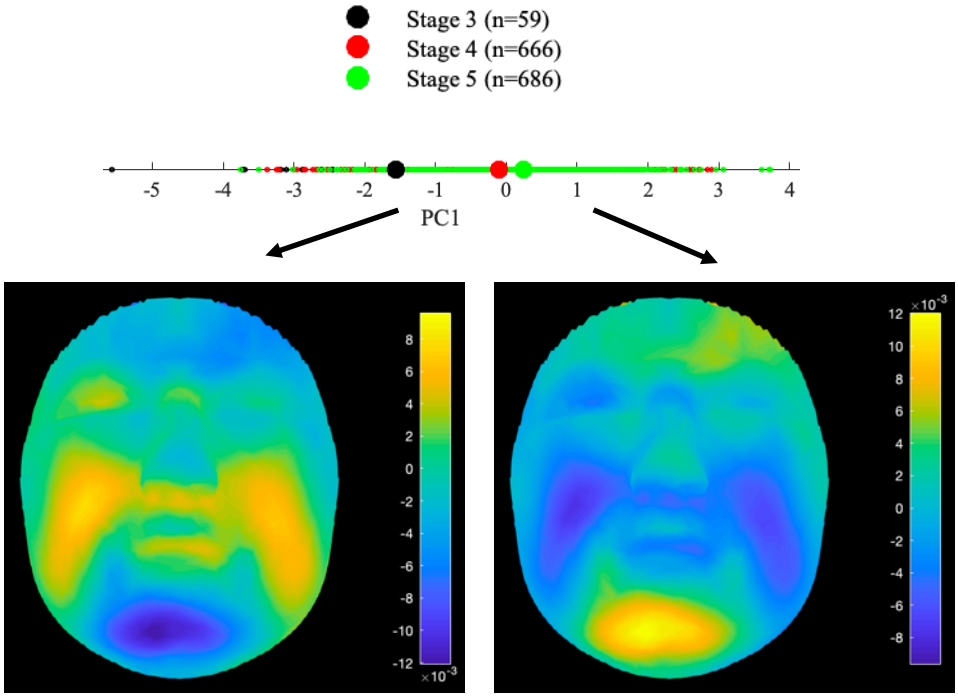


Figure 41: Visualisation of the component scores and facial differences due to pubertal stage (pubic hair) PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.7 Pubertal stage (genital development)

Pubertal stage as determined through self-declared Tanner stage with regards to genital development explained 0.58% of the total variation. The component scores show an ordered separation along PC1, with a separation of stage three (negative mean component score) compared to stage four and five (mean component scores close to origin). Stage three had the smallest sample size. Subjects at stage three were less separated from stage four and five compared to the pubic hair model. Component scores are significantly different (p<0.001). The pattern of facial differences is like the pubic hair model but with the additional difference in nasal shape. The large spread of component scores makes the results more unreliable (Figure 42).

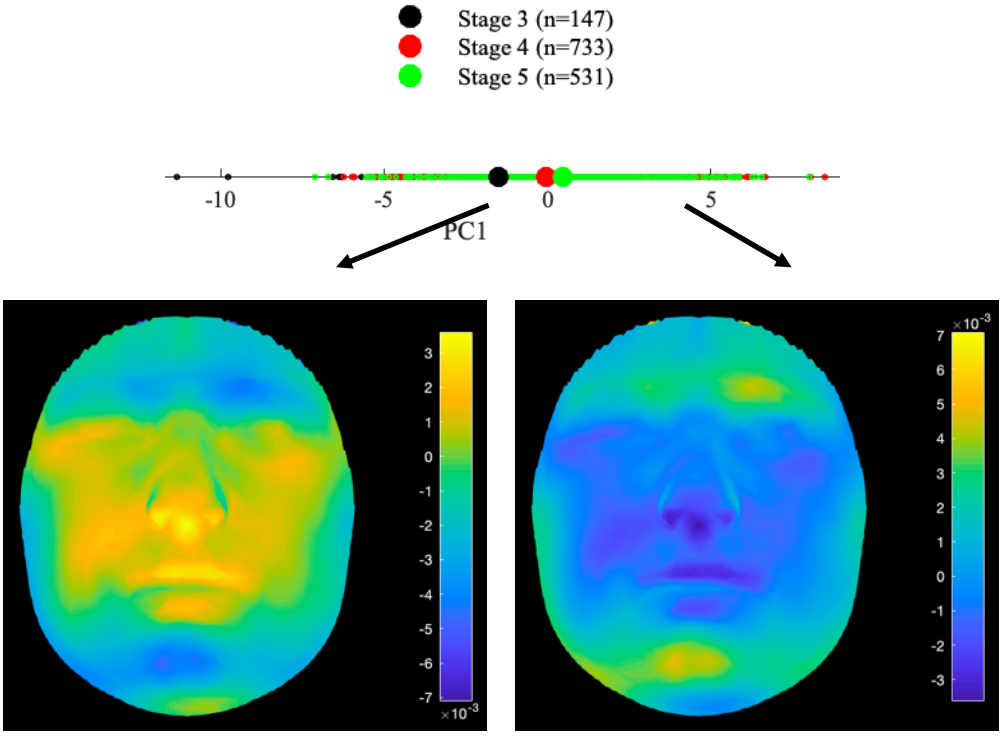


Figure 42: Visualisation of the component scores and facial differences due to pubertal stage (genital development) PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.8 Fasting insulin

Insulin explained 5.83% of the total variation in its 1000 quasi-landmark mPCA model. The component scores showed an ordered pattern along PC1 until the two highest levels of insulin. The mean component scores of the groups with less than 25 iu/ml of fasting insulin showed a progressive increase from negative scores to positive scores. The two groups with the highest levels of fasting insulin did not follow the ordered pattern. These groups had the smallest sample sizes, although do not appear to be disproportionally separated from the other groups. Along PC2, again there is an ordered pattern (close to the origin) except for the highest insulin group which produced negative component scores. This group appears to be disproportionally separated and may therefore be due to the reduced sample size. This, alongside, a lack of a consistent hierarchical pattern makes it more difficult to interpret the scores with confidence. However, the differences for component scores were significant (p<0.001). The facial differences show a similar pattern to those seen for sex, height and BMI, with those with lower insulin levels displaying more retrusive cheeks and more prominent midface (Figure 43).

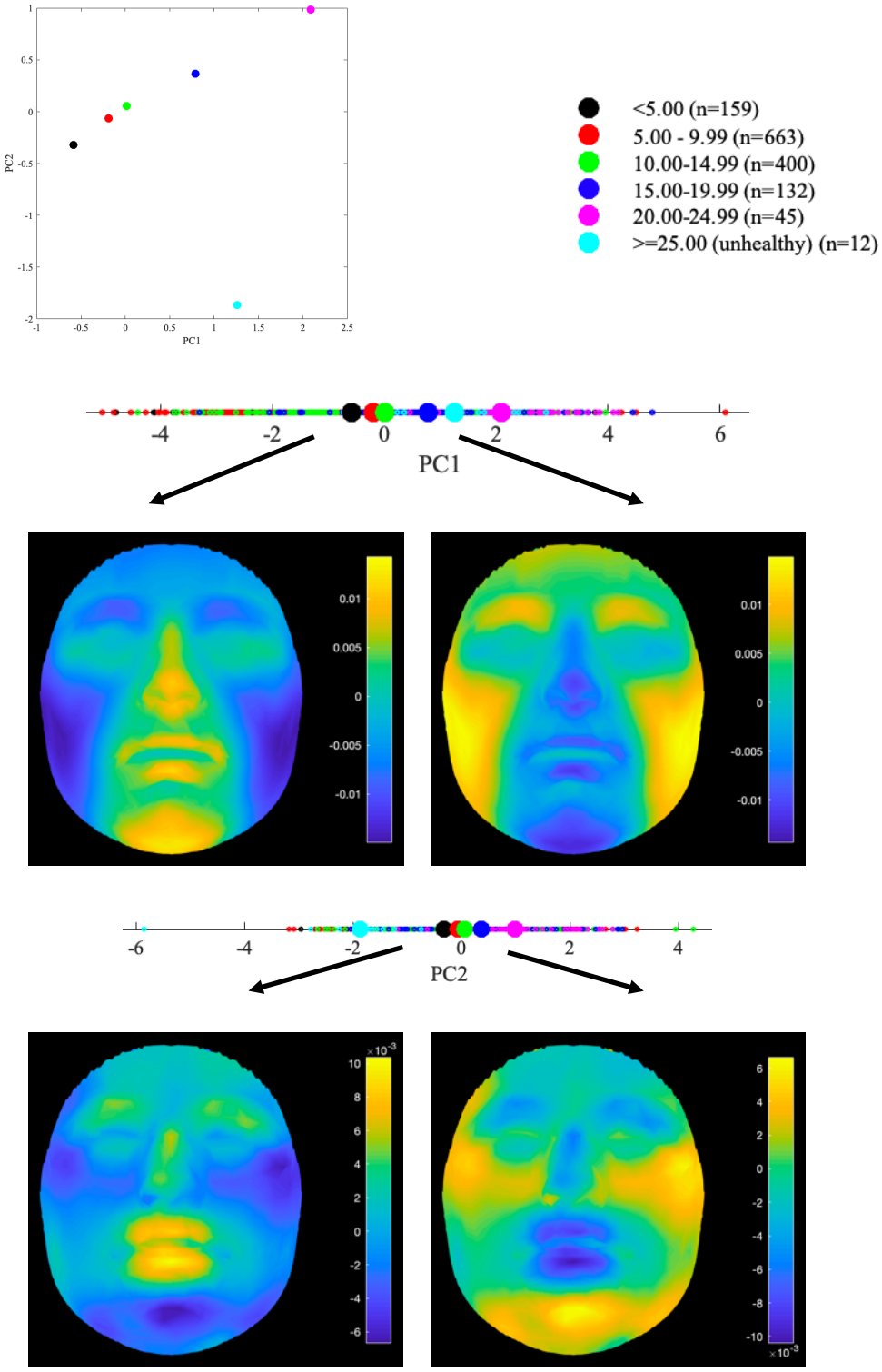


Figure 43: Visualisation of the component scores and facial differences due to insulin PC1 and PC2. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.9 Cholesterol

Cholesterol explained 3.54% of the total variation in its 1000 quasi-landmark model. The component scores showed an ordered pattern, with the subjects with the highest (unhealthy) levels of cholesterol most distinguishable with a negative mean component score. This group also had the smallest sample size. The differences in the component scores were significant (p<0.001). Those with unhealthy levels of cholesterol had larger cheeks with relative midface retrusion. The eyes are also influenced. Those with lower cholesterol levels have a larger brow-ridge as well as nose, lips and chin in comparison to the cheeks, orbits and upper forehead (Figure 44).

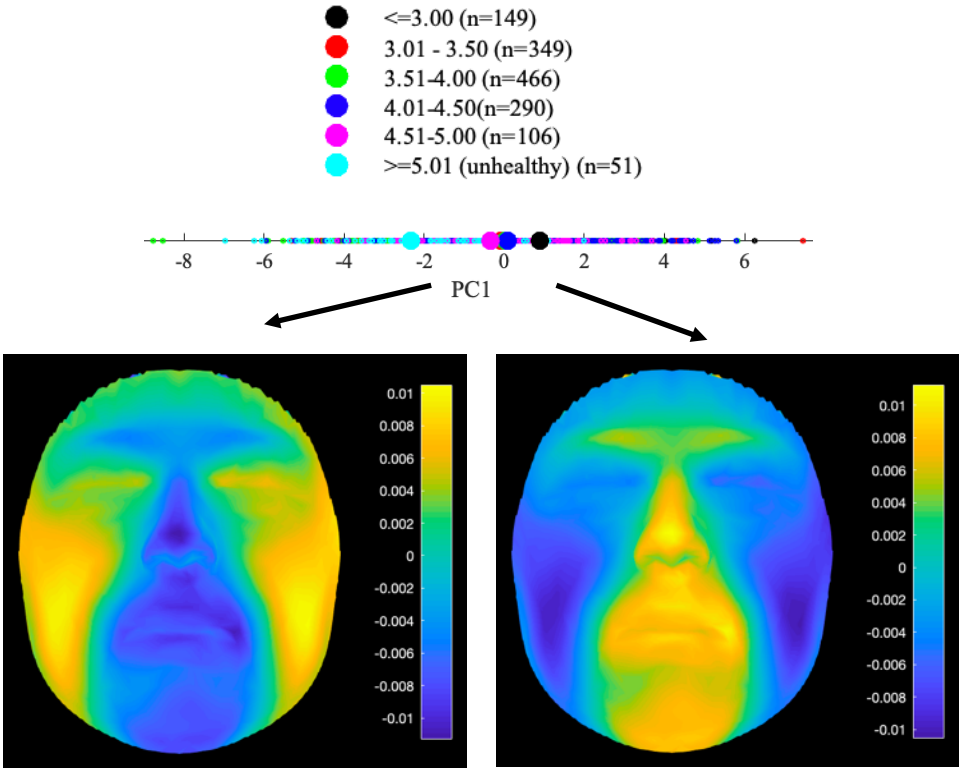


Figure 44: Visualisation of the component scores and facial differences due to cholesterol PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.10 Triglycerides

Triglycerides explained 2.25% of the total variation in its 1000 quasi-landmark model. There appears to be separation of those with higher levels of triglycerides compared to those with less than 1.49mmol/l. These differences were significantly different (p<0.001) for component scores. Those with higher levels of triglycerides had relatively larger cheeks and orbits compared to the midface (Figure 45).

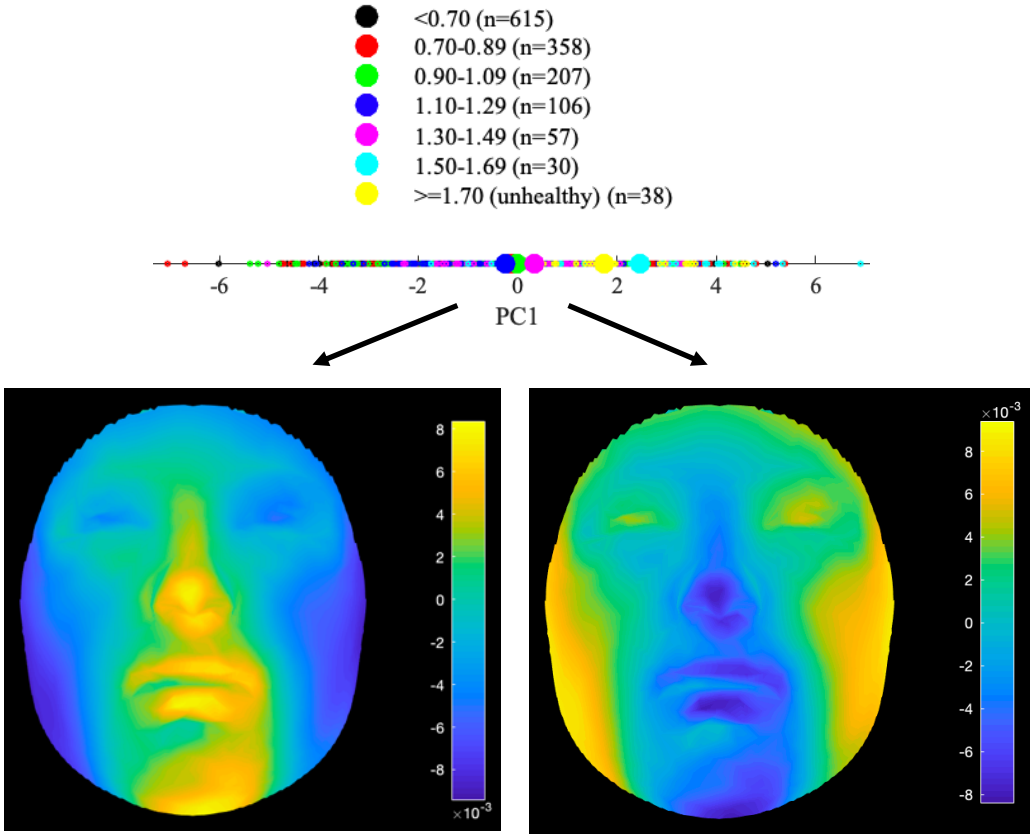


Figure 45: Visualisation of the component scores and facial differences due to triglycerides PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.11 Very Low Density Lipids

VLDL explained 3.54% of the total variation in its 1000 quasi-landmark model. There is a clear separation of the component score mean for the highest levels of VLDL. This group has the smallest sample size. There is also separation of the group with the lowest VLDL in the opposite direction. The differences in the component scores are significant (p<0.001). VLDL appears to negatively influence the shape of the nose, lips and chin in comparison to the cheeks, forehead and eyes (Figure 46).

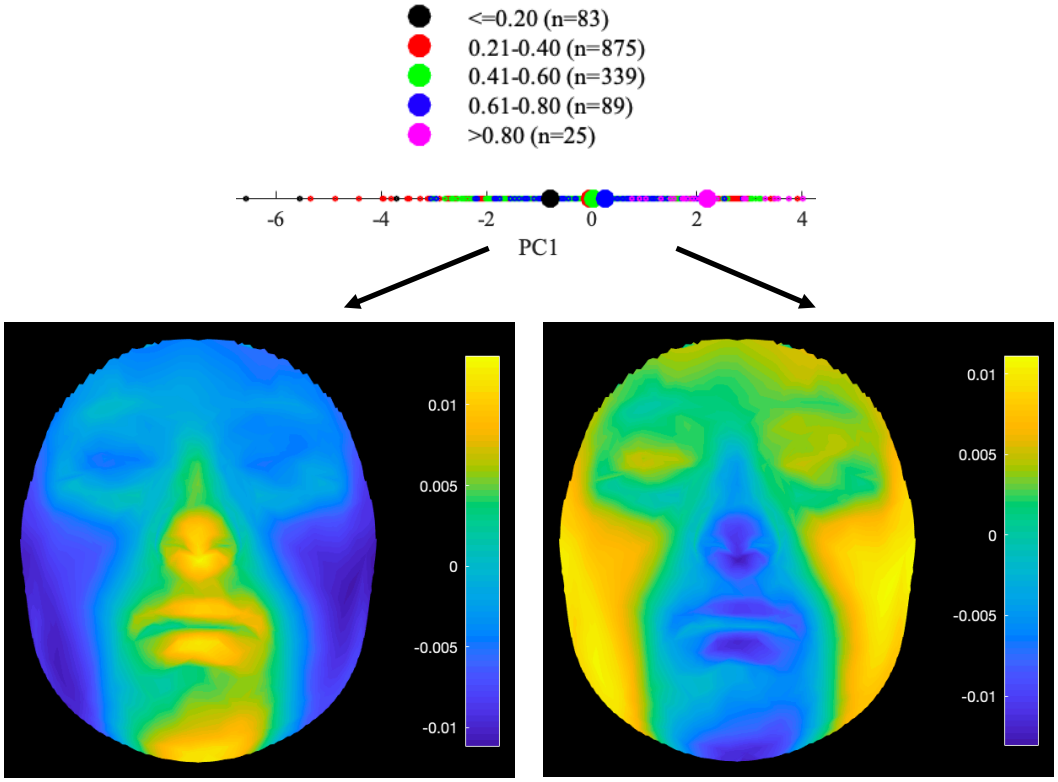


Figure 46: Visualisation of the component scores and facial differences due to VLDL PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.12 Low Density Lipids

LDL explained 2.01% of the total variation in its 1000 quasi-landmark model. There was an ordered separation of the component scores with clear separation of those with unhealthy levels compared to the other groups (positive mean component score in comparison to mean component scores centred around origin or negative). These differences in the component scores are significant (p<0.001). As had been seen with the other variables, the group that was most separated, was also the group with the smallest sample size. Those with unhealthy levels of LDL had larger cheeks and orbits, with a more retrusive midface, including the brow-ridge. In those with lower levels of LDL, the cheeks and orbits were relatively retrusive, whilst the chin, lips, nose and brow-ridge were larger (Figure 47).

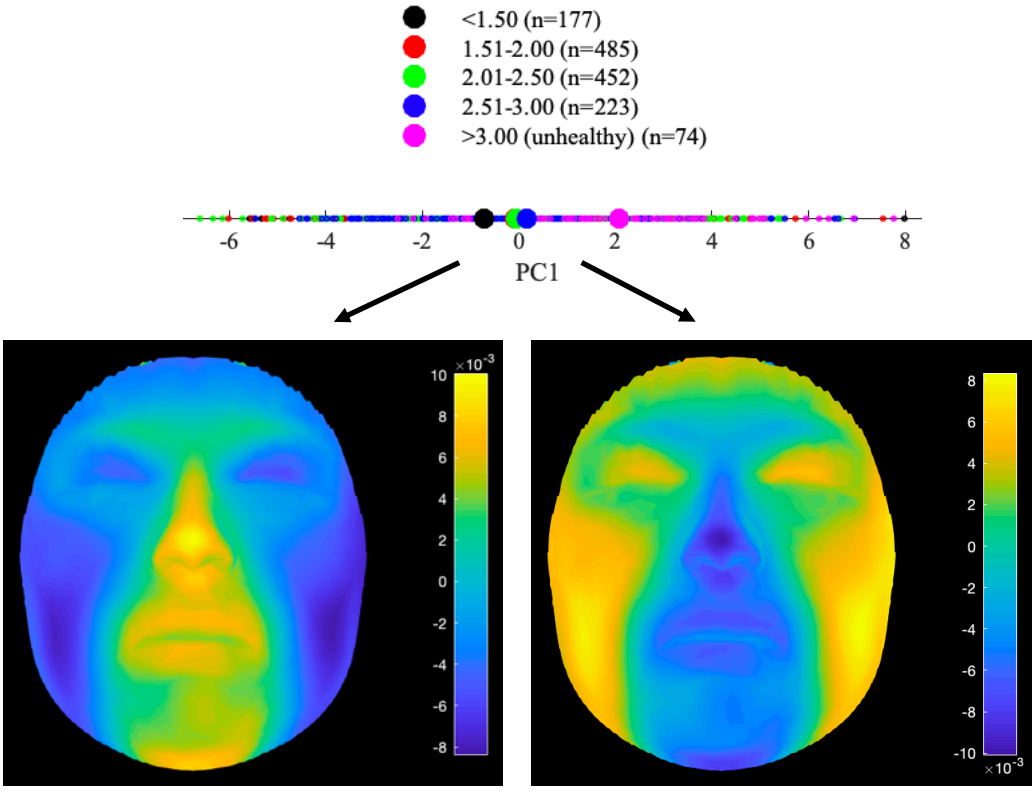


Figure 47: Visualisation of the component scores and facial differences due to LDL PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.13 High Density Lipids

HDL explained 1.58% of the total variation in its 1000 quasi-landmark model. The component scores showed an ordered separation with those with the lowest levels of HDL producing positive mean component scores and those with the highest scores producing negative mean component scores. However, this separation was less marked than that seen for the other variables. This separation was significant for the component scores (p<0.001). The group with the smallest sample size was not disproportionately separated from the other groups. Higher HDL levels influenced the nasal tip, upper zygomatic region, and the infra-orbital region. The area of the cheek influenced is different to that seen in the other variables as it appears to be isolated from the buccal fat pad. The brow-ridge appears to be influenced in a negative direction along either the x-, y- or z- axis (Figure 48).

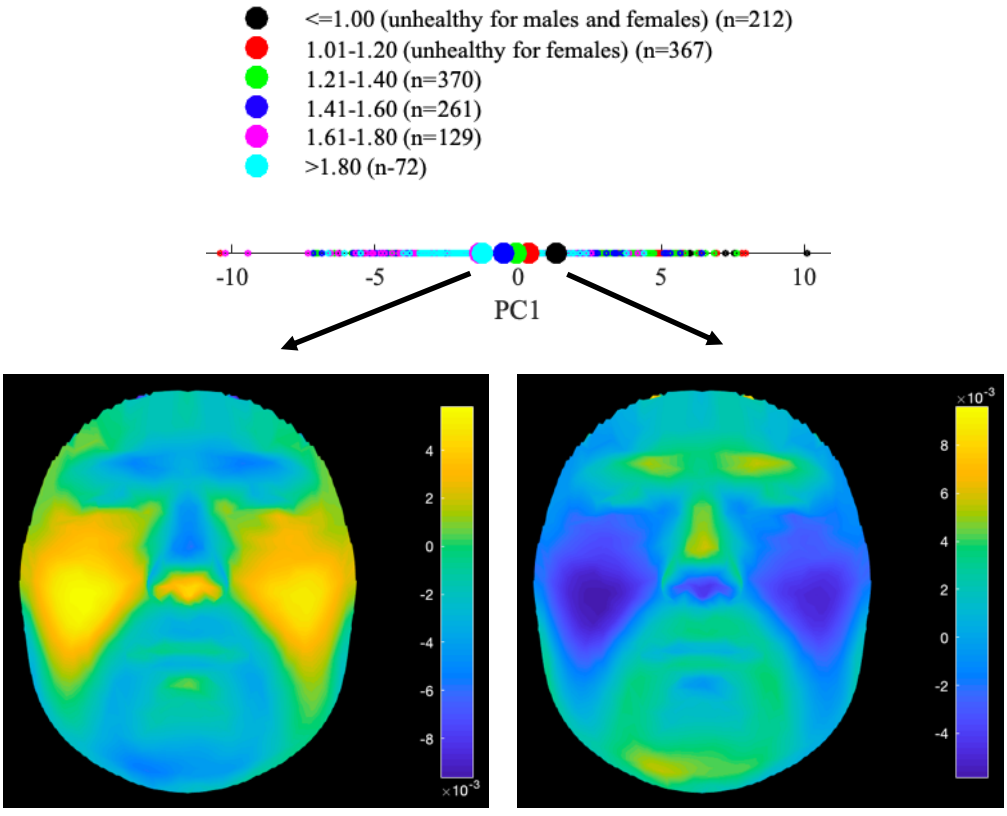


Figure 48: Visualisation of the component scores and facial differences due to HDL PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.14 Glucose

Glucose explained 1.69% of the total variation in its 1000 quasi-landmark model. There was very little separation of the largest groups. However, there is clear separation of the component score means which represents the subjects with the lowest glucose levels. This group has substantially fewer subjects. This separation was statistically significant (p<0.001) for the component scores. Lower levels of glucose appear to influence the upper forehead shape as well as the orbits. The area around the mouth appears to be relatively retrusive or narrower (Figure 49).

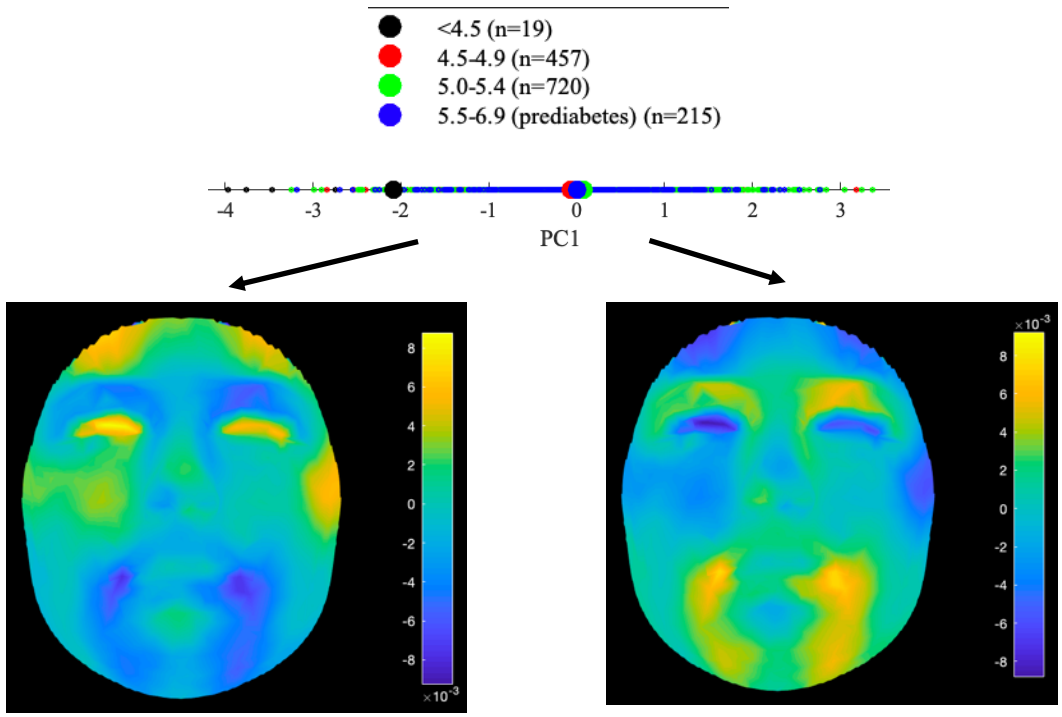


Figure 49: Visualisation of the component scores and facial differences due to glucose PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.15 Atopy

Atopy explained 0.14% of the total variation in its 1000 quasi-landmark model. This was the least of all the models. There was clear separation of the component scores which was statistically significant (p<0.001). The size of the nose, infra-orbital regions, brow-ridge, and mandibular region appear to be relatively positively influenced along the *x*-, *y*- or *z*- axis with atopy, whilst the maxillary region and both lips appear negatively influenced. There is also a suggestion of asymmetry with the left side more affected than the right (Figure 50).

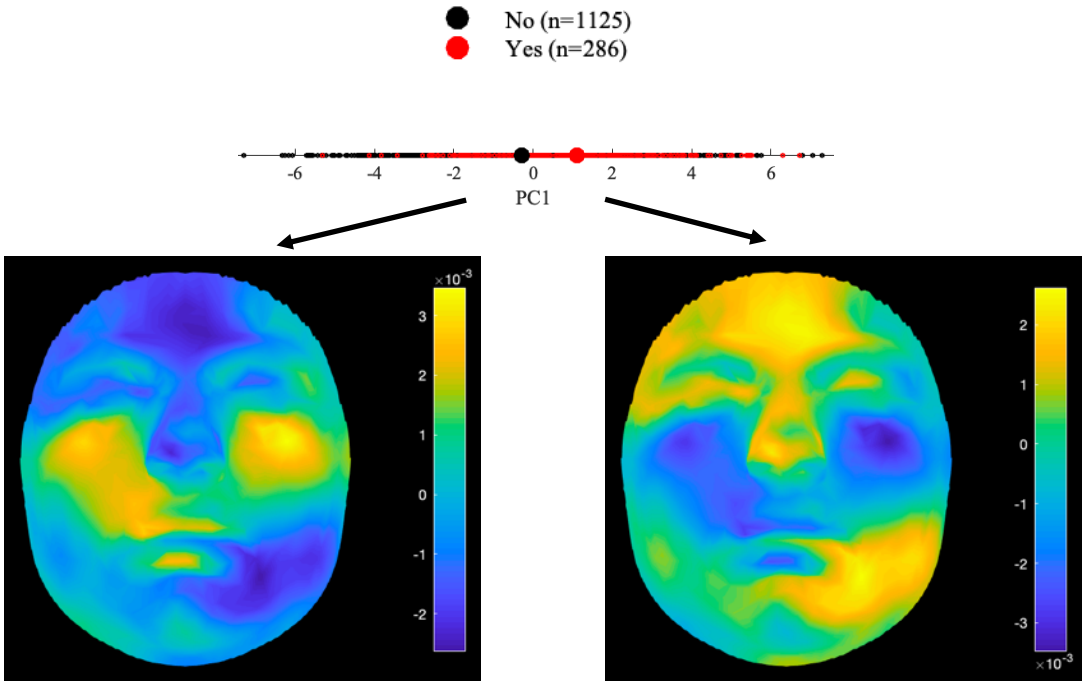


Figure 50: Visualisation of the component scores and facial differences due to atopy PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.16 Hay fever

Hay fever explained 0.35% of the total variation in its 1000 quasi-landmark model. There was clear separation of the component scores, which was statistically significant (p<0.001). The shape of the nasal tip appears to be influenced by hay fever and the opposite direction to the other facial structures. The maxillary region and chin are more negatively influenced (Figure 51).

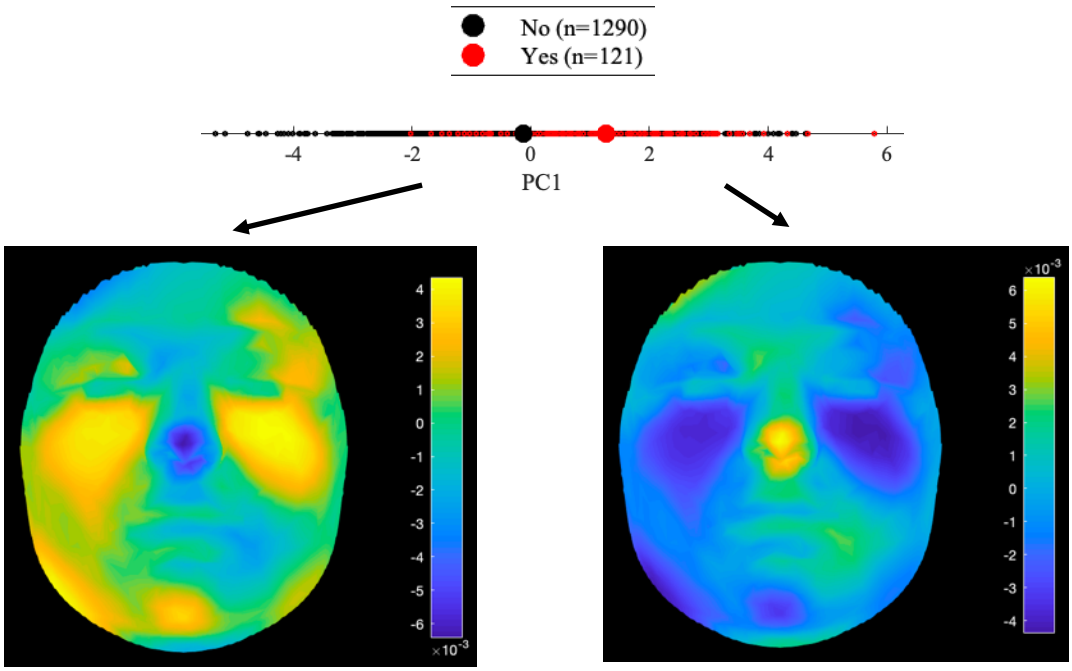


Figure 51: Visualisation of the component scores and facial differences due to hay fever PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.17 Asthma (0-3.5 years old)

Asthma in early childhood explained 0.93% of the total variation in its 1000 quasi-landmark model. There is almost no separation of the subjects who never wheezed and who wheezed at six months only. There is clear separation of those that wheezed at both six months and 42 months. However, as has been seen with other variables, this group had the smallest sample size. The differences between the component scores were significant (p<0.001). Wheezing at both six months and 42 months appears to influence either chin shape or lower facial height (constrained by the scaling procedures). The lower lip is also positively influenced either along the *x*-, *y*- or *z*-axes. The nose and brow-ridge are influenced in the opposite direction (Figure 52).

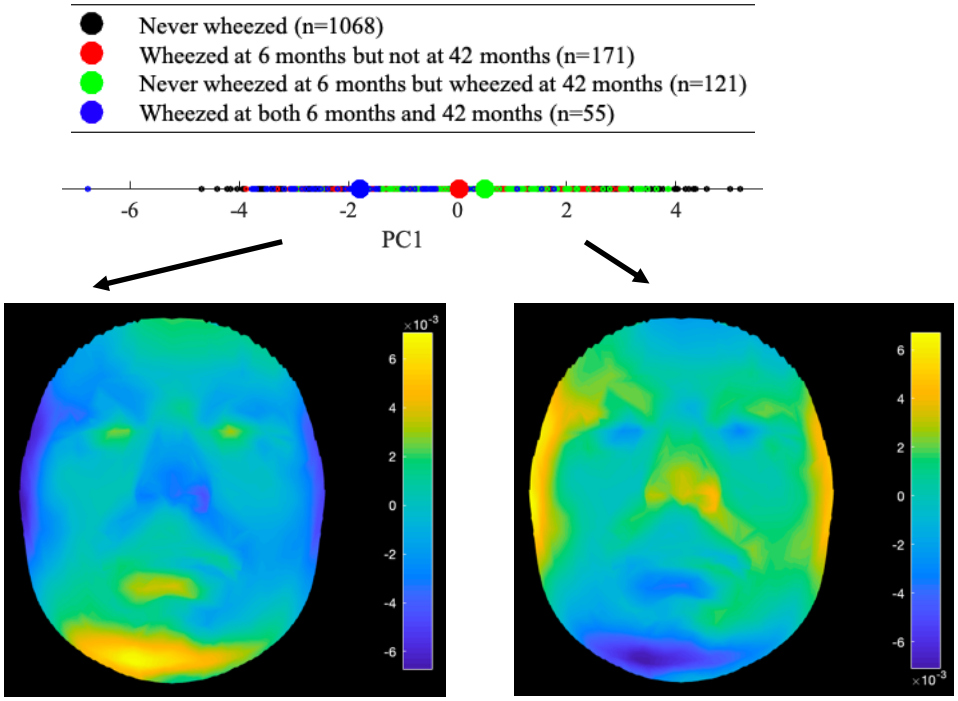


Figure 52: Visualisation of the component scores and facial differences due to asthma (0-3.5 years old) PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.18 Asthma (7.5 years old)

Asthma at 7.5 years old explained 0.40% of the total variation. There is clear separation of the mean component scores for each group. However, there is a large range of scores. The differences between the component scores were significant (p<0.001). Asthma at 7.5 years old appears to show a similar impact to the adolescent facial shape but to a greater extent than that seen at 3.5 years old. The mandibular shape appears to be influenced, but the lower facial height does not appear to be affected as the area directly below the chin does not change from the mean shape. The infra-orbital region is also positively affected in either the *x-, y-* or *z-* axis, whilst the upper lip and maxillary region is influenced in the opposite direction suggesting a reduced midface height or class III malocclusion (Figure 53).

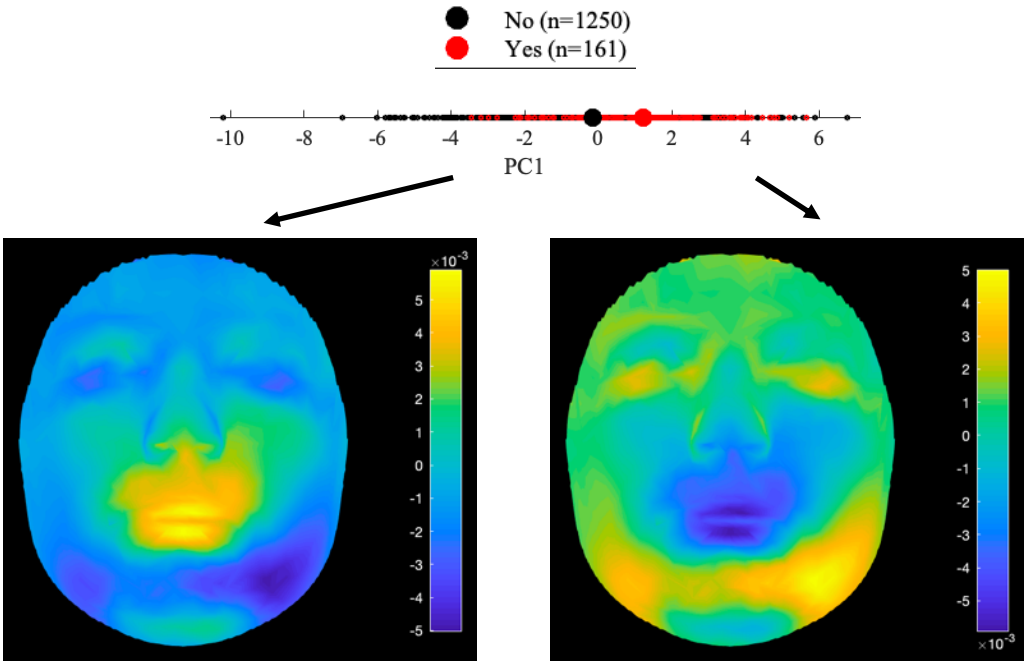


Figure 53: Visualisation of the component scores and facial differences due to asthma (7.5 years old) PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.19 Maternal smoking before pregnancy

Maternal smoking before pregnancy explained 1.53% of the total variation in its 1000 quasi-landmark model. The component scores along PC1, PC2 and PC3 show separation of the mean scores of the groups but this does not appear to be ordered in nature. However, the differences in the component scores were significant (p<0.001). As there does not appear to be any ordered interpretation of the component scores, it is unwise to interpret the facial shape differences at this stage (Figure 54).

5.5.20 Maternal smoking during the 1st trimester

Maternal smoking during the 1st trimester explained 2.51% of the total variation in its 1000 quasi-landmark model. There is clear separation of the component scores of the highest smokers compared to the subjects whose mothers smoked along PC1 with the group sample sizes smallest in the highest smoking group. The differences along PC2 were not ordered, with those whose mothers smoked 15-19 cigarettes per day most differentiable from the other groups. The differences in the component scores were significant (p<0.001). There is a suggestion at PC1 of increased facial proportions with high levels of maternal smoking, with the exception of the nose which appears smaller. PC2 suggests an element of asymmetry at the mandible but this should be interpreted with caution (Figure 55).

5.5.21 Maternal smoking during the 2nd trimester

Maternal smoking during the 2nd trimester explained 3.74% of the total variation in its 1000 quasi-landmark model. The component scores do not show an ordered pattern along PC1 and should therefore be interpreted with caution. The component scores along PC2 do show an ordered pattern with the groups of subjects whose mothers smoked more heavily during pregnancy most differentiable. These groups also have the smallest size. The differences in the component scores were significant (p<0.001). PC1 suggests that pronasale and the forehead may between more prominent with increased maternal smoking, in contrast to the nasal bridge, eyes and lower border of the mandible. PC2 suggests that maternal smoking during the 2nd trimester could influence nasal shape in a negative direction along either *x-, y-* or *z-* axis, whilst the cheeks and orbits are influence in the opposite direction (Figure 56).

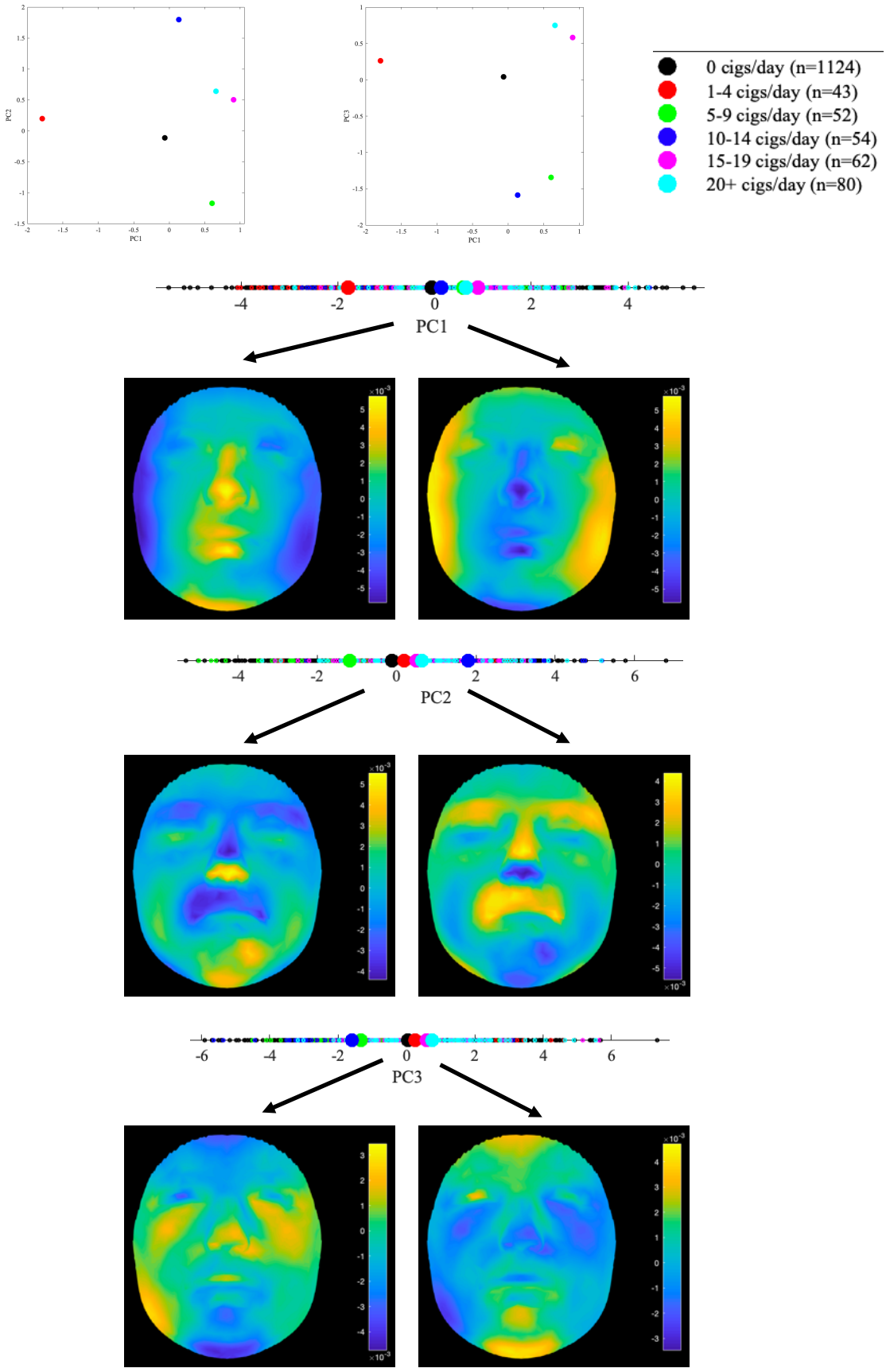


Figure 54: Visualisation of the component scores and facial differences due to maternal smoking before pregnancy PC1, PC2 and PC3. Yellow = larger/ more prominent; blue = smaller, more retrusive.

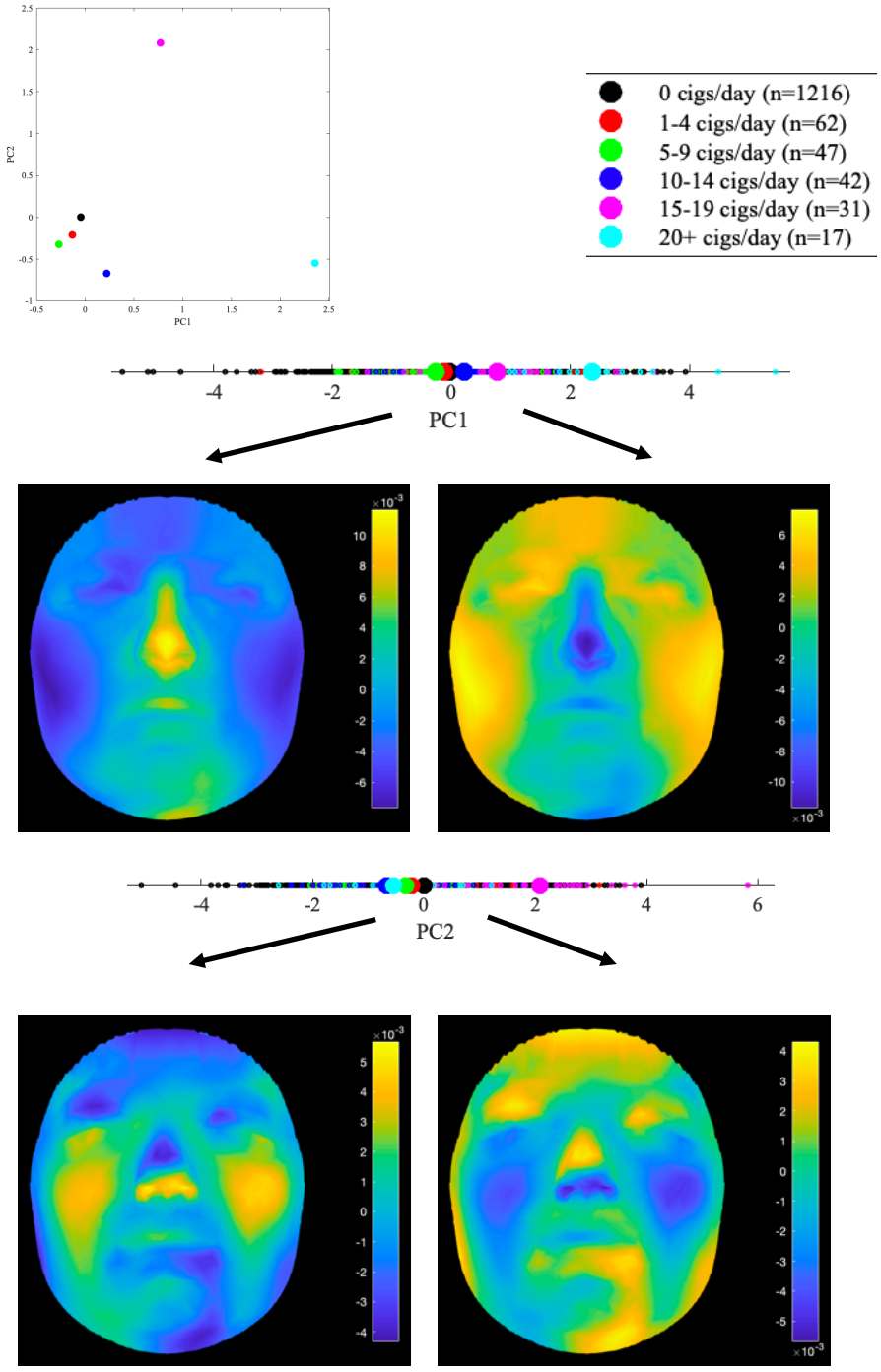


Figure 55: Visualisation of the component scores and facial differences due to maternal smoking during the 1st trimester PC1 and PC2. Yellow = larger/ more prominent; blue = smaller, more retrusive.

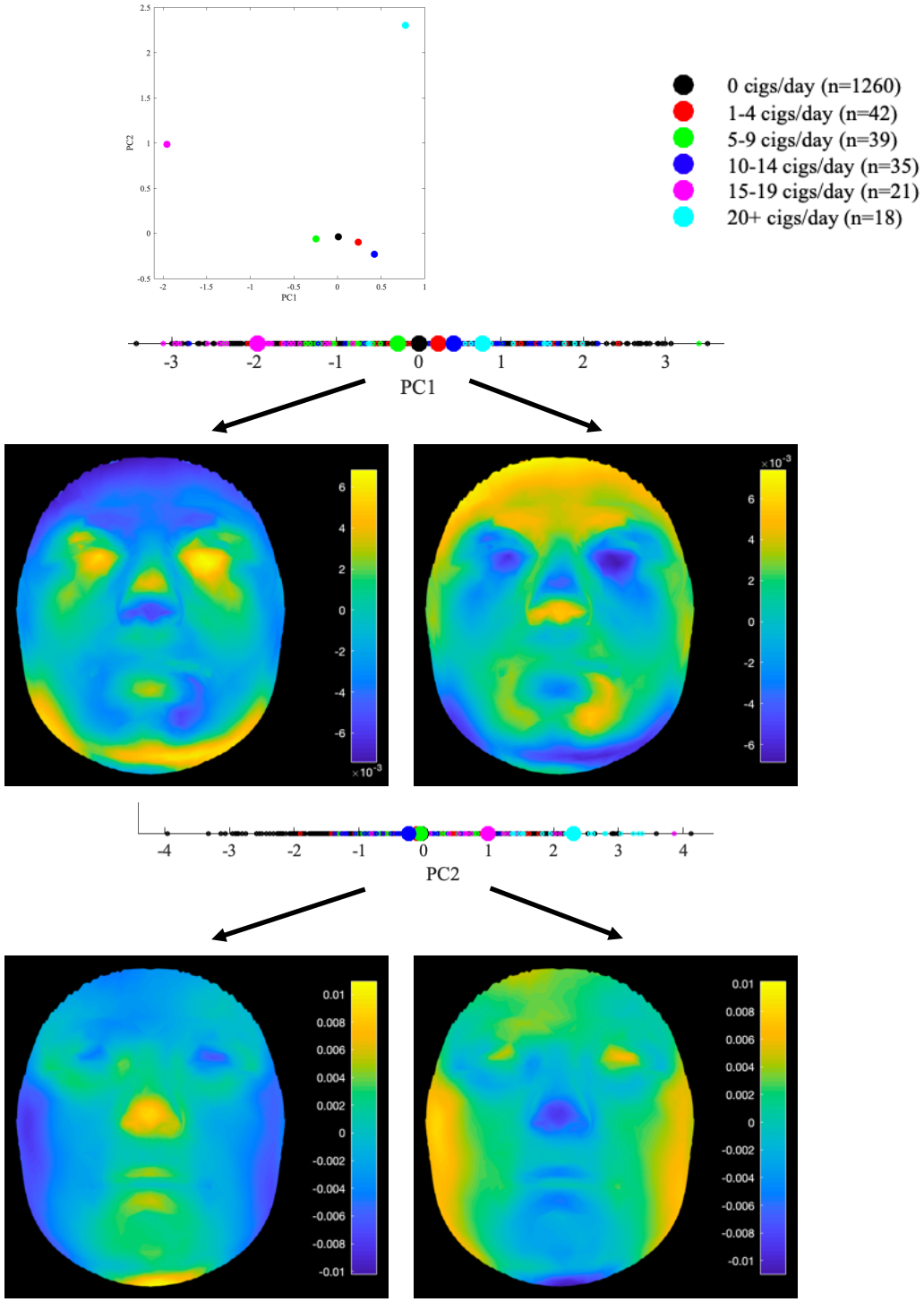


Figure 56: Visualisation of the component scores and facial differences due to maternal smoking during the 2nd trimester PC1 and PC2. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.22 Maternal alcohol consumption before pregnancy

Maternal alcohol consumption before pregnancy explained 1.56% of the total variation in its 1000 quasi-landmark model. Those whose mothers did not drink alcohol before pregnancy are clearly distinguishable from those whose mothers drank any level of alcohol before pregnancy. Again, the most distinguishable group had the smallest sample size. The differences between the component scores were significant (p<0.001). The visualisations suggest that drinking any level of alcohol before pregnancy could reduce the shape of the nasal tip and bridge in an x-, y- or z- direction and the underside of the chin. There may also be an influence on the shape of the forehead in a positive direction along one or more of these axes as well as the infra-orbital regions (Figure 57).

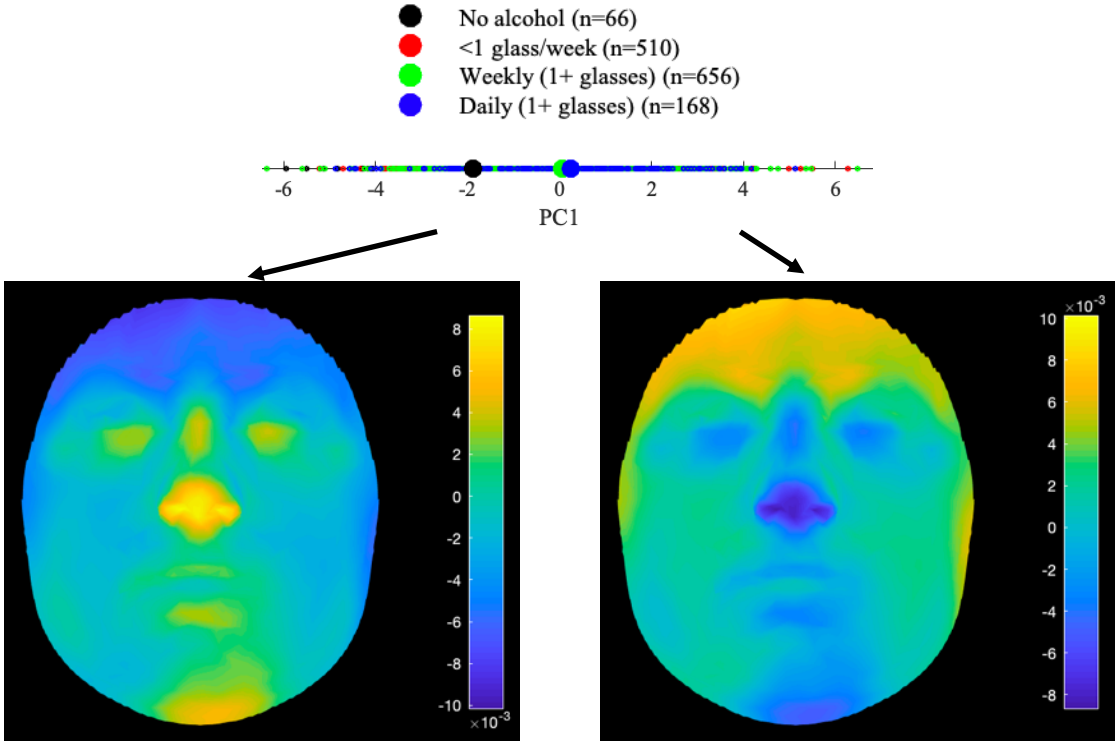


Figure 57: Visualisation of the component scores and facial differences due to maternal alcohol consumption before pregnancy PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.23 Maternal alcohol consumption during the 1st trimester

Maternal alcohol consumption during the 1st trimester explained 2.62% of the total variation in its 1000 quasi-landmark model. There was clear separation of the component scores for those whose mothers drank daily during the 1st trimester of pregnancy compared to those whose mothers did not drink or consumed alcohol on a weekly or less than weekly basis. However, this group had the smallest sample size. The differences in the component scores were significant (p<0.001). Daily maternal alcohol consumption during the 1st trimester of pregnancy appears to negatively influence the shape of the nose in either *x-, y-* or *z-* directions. The shape of the chin, or perhaps the lower facial height (constrained by the scaling procedures) also appears to be negatively affected. In contrast, the forehead, supra-orbital region, cheeks and lips appear to be influenced in a positive direction along either *x-, y-* or *z-* axes (Figure 58).

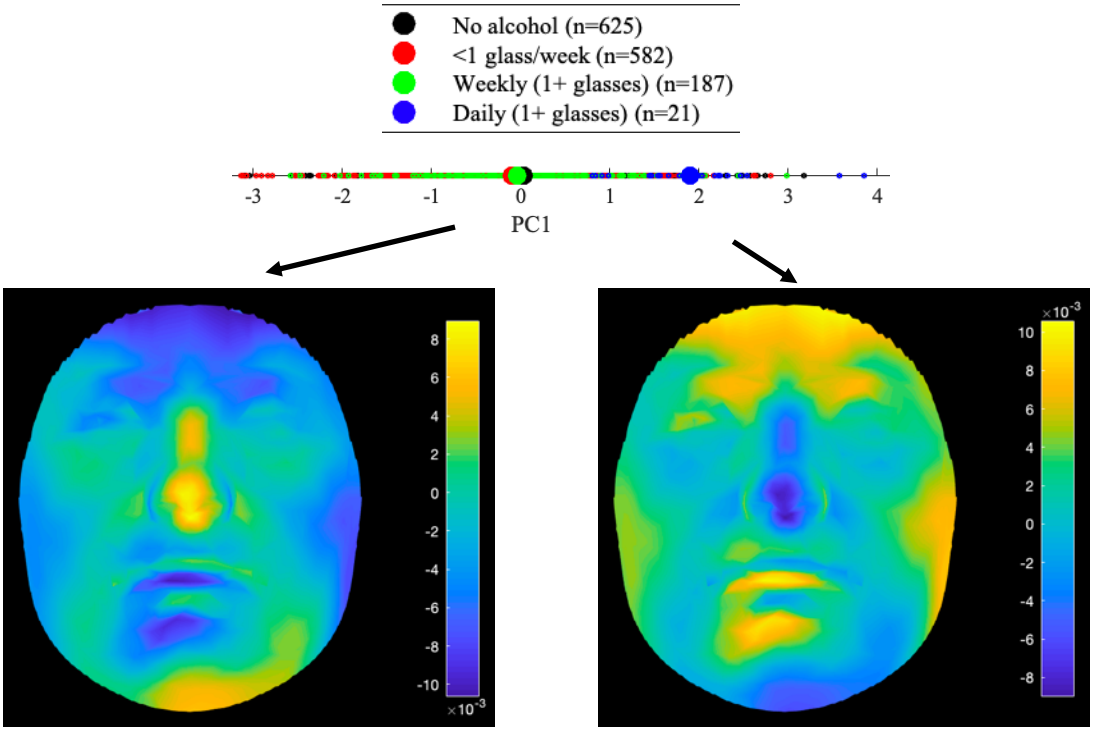


Figure 58: Visualisation of the component scores and facial differences due to maternal alcohol consumption during the 1st trimester PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.5.24 Maternal alcohol consumption during the 2nd trimester

Maternal alcohol consumption during the 2nd trimester explained 1.60% of the total variation in its 1000 quasi-landmark model. Again, those whose mothers drank daily during the 2nd trimester of pregnancy were most differentiable and this group had the smallest sample size. The differences in the component scores were significant (p<0.001). Daily maternal alcohol consumption during the 2nd trimester appears to include the chin and forehead in a positive direction, with an element of asymmetry. The nose is influenced in a negative manner along the bridge and pronasale, it may be wider in the transverse axis (Figure 59).

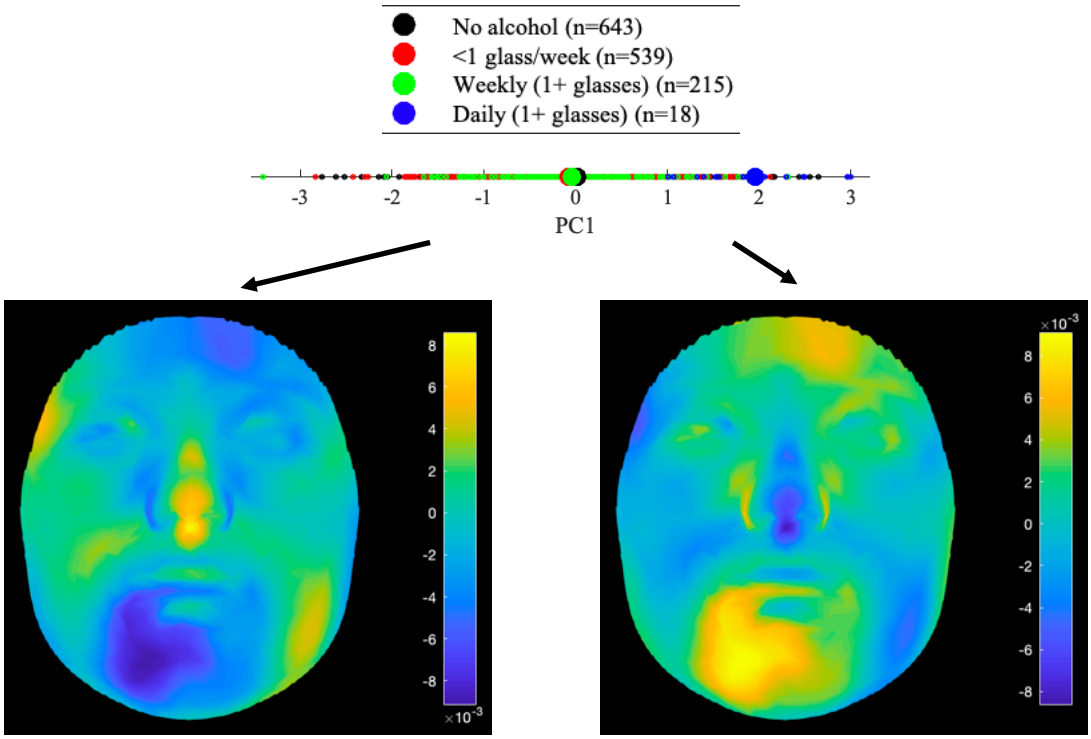


Figure 59: Visualisation of the component scores and facial differences due to maternal alcohol consumption during the 2nd trimester PC1. Yellow = larger/ more prominent; blue = smaller, more retrusive.

5.6 DISCUSSION

As in study one, this discussion focuses on the relative advantages and disadvantages of mPCA for investigating facial shape. Discussion of the influence of each variable is provided in the General Discussion (7.1 Facial differences, p. 165).

mPCA was to be able to maximise the differences due to each variable and provide clear visualisations of the overall differences in the form of scatter plots. mPCA therefore potentially provides a method of determining the influence of variables that have a small, but significant, influence on facial shape. However, interpretation of the component scores must be tentative at this stage, given that in many examples, the group with the smallest sample size was isolated from the other groups. It has been suggested by Bookstein (2019) that in between-groups PCA, which is a related technique to mPCA, group component scores can be artificially isolated to the end of an axis when a group sample size is much smaller than the others. This phenomenon may also occur for mPCA, thus it is challenging to ascertain whether the differences in the component score means are due to true differences or whether this is an artefact due to differing group sample sizes.

Perhaps the most advantageous use of mPCA that has been demonstrated in this study is the ability to compare the relative importance of each variable. This process takes seconds and could therefore easily be added into future data exploration phases of data analysis. It is important that mPCA is not carried out in isolation as the models are not independent of each other as there is no adjustment for variables unless they are added as separate levels in the model. This is particularly salient in this study as many of the variables were correlated with each other (3.7.4.2 Correlation between variables, p. 66). This is highlighted when comparing the results for sex, height, BMI and insulin. All these variables appear to influence the same facial features, to differing degrees (Figure 60). Given that sex was correlated with height and BMI with fasting insulin (3.7.4.2 Correlation between variables, p. 66), it is not possible to ascertain how much of the differences in facial features are due to the variables in isolation using the two-level models alone.

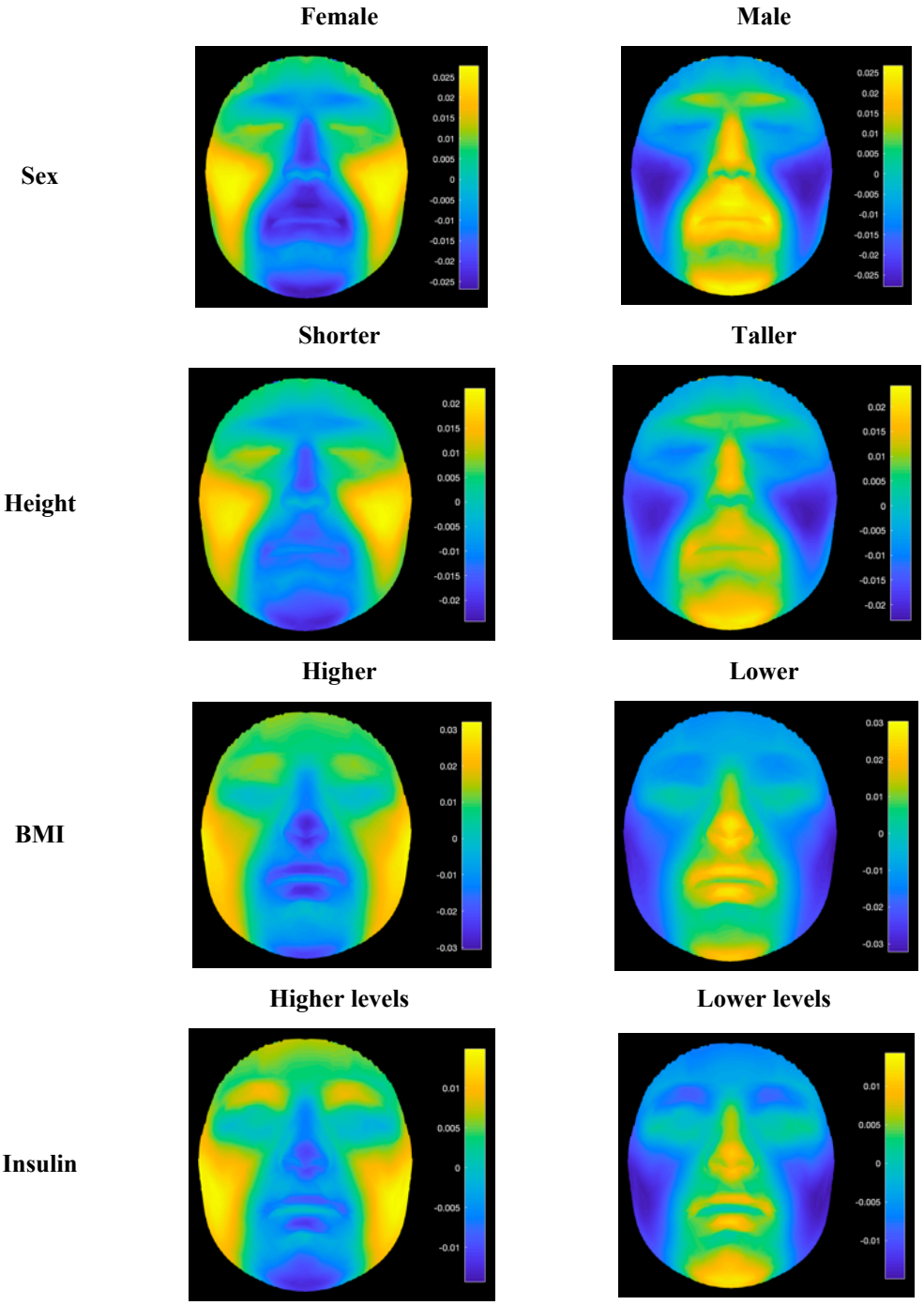


Figure 60: A comparison of the results for sex, height, BMI and insulin. mPCA results show very similar differences in the facial shapes but on subtly different scales. As these results were gained via separate two-level models, and mPCA does not correct for variables that are not in the model, the true influence of each variable is unclear. Yellow = larger/ more prominent; blue = smaller, more retrusive.

Unfortunately, the sample constraints required for each group limited further exploration of BMI and insulin using an mPCA model with additional levels. When considered using their current groupings, there were groups with no subjects (Figure 61). mPCA cannot currently support groups with zero subjects. A four-level mPCA model was possible to assess the influence of maternal smoking, maternal alcohol consumption, sex and within-group variation. However, due to sample size constraints, discretisation was limited to exposure and non-exposure only during the first and second trimesters (Galloway et al. 2020). This limited the information available from the analysis. It was felt that exploring fasting insulin and BMI further using different groupings in the mPCA model would also limit the impact of the analysis, particularly as pathological and non-pathological levels of insulin would not be separated if groups were merged. Alternative methods were subsequently investigated that do not require discretisation of the variables.

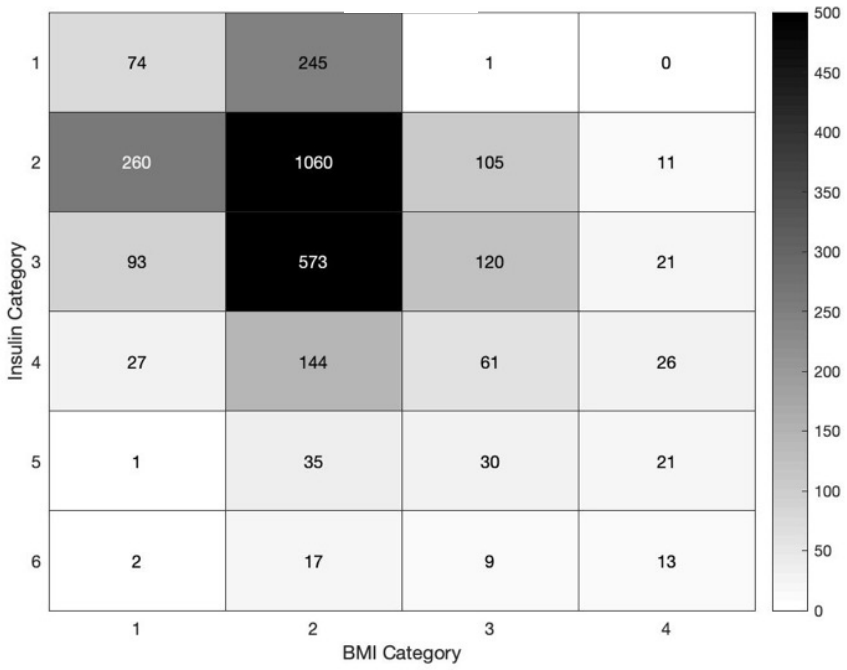


Figure 61: Heatmaps showing sample sizes for groupings if insulin and BMI added as separate levels in a 3-level mPCA model. Each square of the heatmap represents the number of subjects in each potential group in the mPCA model.

PLSR provides an alternative technique that is capable of handling both continuous and categorical variables (Shrimpton et al. 2014). The issue of discretisation is therefore overcome. However, PLSR models are computationally expensive and, like all models, are prone to overfitting if excessive numbers of variables are added to the model (Tabachnick and Fidell 1996). Using mPCA as a variable selection tool prior to PLSR by exploring variables that meet a predetermined threshold with regards to percentage of the total variation explained in their respective mPCA model, may therefore be of benefit.

5.7 SUMMARY

A summary of the inferential findings from this study are found in Table 14. This study has introduced several disadvantages of using mPCA in facial shape research. Firstly, the requirement to categorise the variables can introduce the issue of imbalanced group sizes and the potential for groups with zero subjects. This in turns limits the number of levels possible and the possibility of adjusting for variables in the analysis. There is also uncertainty around whether group mean component scores are artificially placed at the end of an axis, thus it is unwise to trust mPCA results in isolation.

Using a further analysis to confirm the results gained from the mPCA models would therefore be worthwhile. A method previously used within this field is PLSR (Shrimpton et al. 2014). However, difficulty can arise in deciding which variables to add to parsimonious PLSR models. mPCA may provide a useful tool for variable selection by using the percentage of the total variation explained by each of the variables in their respective mPCA models as a threshold for inclusion in a PLSR model. This concept will be explored in study three.

Table 14: Summary of results from study two

Variable	Test	% Total variation explained	*P*-value	Threshold for significance	Null hypothesis
Sex	mPCA + ANOVA	17.31%	<0.001	0.002	Reject
Age	mPCA + MANOVA	1.18%	>0.05	0.002	Accept
Height	mPCA + MANOVA	11.05%	<0.001	0.002	Reject
BMI	mPCA + MANOVA	17.29%	<0.001	0.002	Reject
Pubertal stage (hair)	mPCA + MANOVA	1.76%	<0.001	0.002	Reject
Pubertal stage (genitals)	mPCA + MANOVA	0.58%	<0.001	0.002	Reject
Cholesterol	mPCA + MANOVA	3.55%	<0.001	0.002	Reject
Triglycerides	mPCA + MANOVA	2.25%	<0.001	0.002	Reject
VLDL	mPCA + MANOVA	3.54%	<0.001	0.002	Reject
LDL	mPCA + MANOVA	2.01%	<0.001	0.002	Reject
HDL	mPCA + MANOVA	1.58%	<0.001	0.002	Reject
Glucose	mPCA + MANOVA	1.69%	<0.001	0.002	Reject
Insulin	mPCA + MANOVA	5.83%	<0.001	0.002	Reject
Asthma (0-42 months)	mPCA + MANOVA	0.93%	<0.001	0.002	Reject
Asthma (91 months)	mPCA + ANOVA	0.40%	<0.001	0.002	Reject
Atopy	mPCA + ANOVA	0.15%	<0.001	0.002	Reject
Hay fever	mPCA + ANOVA	0.35%	<0.001	0.002	Reject
Smoking before pregnancy	mPCA + MANOVA	1.53%	<0.001	0.002	Reject
Smoking 1st trimester	mPCA + MANOVA	2.51%	<0.001	0.002	Reject
Smoking 2nd trimester	mPCA + MANOVA	3.74%	<0.001	0.002	Reject
Alcohol before pregnancy	mPCA + MANOVA	1.56%	<0.001	0.002	Reject
Alcohol 1st trimester	mPCA + MANOVA	2.62%	<0.001	0.002	Reject
Alcohol 2nd trimester	mPCA + MANOVA	1.60%	<0.001	0.002	Reject

6 STUDY 3: USING MULTILEVEL PRINCIPAL COMPONENT ANALYSIS AS A VARIABLE SELECTION TOOL PRIOR TO PLSR TO INVESTIGATE THE INFLUENCE OF MULTIPLE CATEGORICAL AND CONTINUOUS VARIABLES ON THE FACIAL SHAPE OF ENGLISH ADOLESCENTS

6.1 INTRODUCTION

As discussed in study two, there are disadvantages to using mPCA in isolation. These are emphasised when assessing continuous variables and when group sample sizes are imbalanced or place limits on the number of levels possible in the model. An alternative technique which has been used previously in facial shape research that is capable of handling both continuous and categorical variables is PLSR (Shrimpton et al. 2014). As PLSR can adjust for variables in the model, it has the potential to provide more robust results than mPCA. However, this technique is computationally expensive and is at risk of overfitting as the number of variables included in the model increases. mPCA may provide a method for determining which variables are most likely to influence facial shape and are therefore most worthwhile including in a PLSR model. This concept will be explored whilst the use of an alternative technique provides a comparison to the results of mPCA.

6.2 AIMS

- To explore the use of mPCA as a variable selection tool prior to PLSR.
- To compare the results of mPCA and PLSR.
- To determine the influence of sex, height, BMI, metabolic factors, maternal smoking during pregnancy and maternal alcohol consumption during pregnancy on the facial shape of English adolescents using PLSR.

6.3 NULL HYPOTHESES

- Sex is not associated with the facial shape of English 14 to 16 year olds.
- Height is not associated with the facial shape of English 14 to 16 year olds.
- BMI (and weight) are not associated with the facial shape of English 14 to 16 year olds.
- Metabolic factors (fasting insulin, cholesterol, triglycerides, LDL and VLDL) are not associated with the facial shape of English 14 to 16 year olds.
- Maternal smoking during the 1st or 2nd trimesters is not associated with the facial shape of English to 14 to 16 year olds.
- Maternal alcohol consumption during the 1st trimester is not associated with the facial shape of English 14 to 16 year olds.

6.4 METHODOLOGY OVERVIEW

An overview of the methodology used in this study is provided below.

6.4.1 3D facial scan acquisition, processing and landmarking

The 3D facial scans were acquired and processed as documented previously. GPA was conducted on the 3D facial scans. Scaling of the images removed the influence of size, thus isolating the influence of shape (3.4 Facial scan processing, p. 53). The analyses used 7160 quasi-landmarks placed automatically using a publicly available MATLAB algorithm (3.5 Landmarking, p. 52).

6.4.2 Variables

The outcome variables were the 7160 quasi-landmarks. The predictor variables were selected from those assessed in study two (3.6 Variables, p. 56). Two thresholds were explored based on the total variation explained by each variable in study two (5.5.1 Percentage of total variation, p. 108).

1) Variables that explained greater than 2% of the total variation in their respective mPCA models in study two: Sex, height, BMI (checked with weight also), fasting

insulin, cholesterol, triglycerides VLDL, maternal smoking during the first and second trimesters, and maternal alcohol consumption during the first trimester.

2) Variables that explained greater than 5% of the total variation in their respective mPCA models in study two: Sex, height, BMI (checked with weight also) and fasting insulin.

Sex was the only binary variable and was coded as a dummy variable (male = 0, female = 1). Height, BMI, insulin, cholesterol, triglycerides and VLDL were naturally continuous and used in the model as such. Maternal smoking and alcohol consumption during pregnancy were collected as ordinal data and were used in the PLSR model in this manner.

6.4.3 Data cleaning and exploration

Outliers were detected in the variables (3.7.3.2 Outliers in variable data, p. 63). However, it is unclear whether these are true outliers or whether they are due to variation in the data. As regression methods are sensitive to outliers, the PLSR models were run with and without outliers for completeness. As VLDL/cholesterol and LDL/triglycerides were strongly correlated (3.7.4.2 Correlation between variables, p. 66), the appropriate models were also repeated without VLDL and LDL. As discretisation is not required for PLSR, the continuous variables were used in their natural form. As the maternal smoking and alcohol data was collected in an ordinal form, they were used in this study as such.

6.4.4 Final sample

The 1411 subjects used in this study were the same as those used in the study two (5.4.4 Final sample, p. 104).

6.4.5 Analyses

PLSR is detailed in 2.9 Analysis of 3D facial landmarks (p. 40). PLSR was carried out in MATLAB R2017b using the PLSHypothesisTests class developed in KU Leuven as part of their imgGenesSoftware. This is based around the MATLAB

function plsregress and is documented in Shrimpton et al. (2014). Permutation methods on the partial R^2 values are included in the PLSHypothesisTests class to provide a *p*-value for each of the variables included in the model. Bonferroni correction was used to account for multiple testing. The facial differences explained by each of the variables were visualised using the 3D viewer class in MATLAB 2017b. The differences in the partial R^2 scores at each landmark and the coefficients attributed to each landmark in the PLSR model were visualised.

Numerous PLSR models were used. These explored the inclusion of variables that explained more than 2% and 5% of the total variation in their respective mPCA models; the inclusion and exclusion of outliers; the inclusion of interaction terms; and the separation of the sexes. The models are summarised in Table 15.

Table 15: PLSR models

Model	Permutations	Outliers included	Sexes included	No. of variables included	Interactions included
1	10,000	Yes	Both	11	0
2	10,000	Yes	Both	9 (VLDL and LDL excluded)	0
3	10,000	No	Both	11	0
4	10,000	No	Both	9 (VLDL and LDL excluded)	0
5	10,000	No	Females	10	0
6	10,000	No	Males	10	0
7	1,000	Yes	Both	4	All possible (2-way)
8	10,000	Yes	Both	4 (Weight not BMI)	All possible (2-way)
8	10,000	Yes	Both	4	Sex * Height Sex * Insulin
9	10,000	Yes	Both	4	0
10	1,000	No	Both	4	All possible (2-way)
11	10,000	No	Both	4 (Weight not BMI)	All possible (2-way)
12	10,000	No	Both	4	Sex * Height Sex * BMI
13	10,000	No	Both	4	Sex * BMI
14	10,000	No	Both	4 (Weight not BMI)	Sex * Height
15	10,000	No	Both	4	0
16	10,000	No	Both	4 (Weight not BMI)	0
17	10,000	No	Females	3	0
18	10,000	No	Males	3	0

6.5 RESULTS

6.5.1 Global *p*-values

The first model was run with outliers included and repeated with the outliers excluded. Given that VLDL and cholesterol, and LDL and triglycerides were strongly correlated using Pearson correlation, the model was re-run with VLDL and LDL excluded. The inclusion of VLDL and LDL did not influence the global *p*-values of the other variables significantly. The variables that resulted in a significant global *p*-values were sex, height, BMI and insulin. The PLSR model with the outliers excluded was repeated for males and females separately. Height and BMI were significant at a 0.0045 level. Insulin was no longer significant when the sexes were assessed separately (Table 16).

Following the results of these models, sex, height, BMI and insulin were explored further. These variables explained more than 5% of the total variation in their respective mPCA models. Initial models, with and without outliers, were run including all the possible interactions. When outliers were included, none of the interactions were significant. The model was therefore repeated without the interactions. In all the models, sex, height, BMI and fasting insulin reached significance at a 0.0125 level. When outliers were excluded, the interaction between sex and BMI reached significance. The model was therefore repeated for the sexes separately. Again, as was seen in the initial models, height and BMI remained significant, but insulin was no longer significant when the sexes were explored separately (Table 17). Finally, as BMI is a construct of weight and height, the model was re-run with weight instead of BMI (sex, height, weight, insulin and their interactions) with and without outliers. This did not appear to influence the resulting significance levels of any of the variables (Table 18).

6.5.2 Global R^2 values

The R^2 values represent model fit (Farnell et al. 2021). The global R^2 values for sex are the highest of all the variables. However, are generally low at 0.045 to 0.047 in the different models. This may be explained since sex explained a relatively low amount of variation in the mPCA model compared to within-group variation. The global R^2

values for height was also low (0.08 to 0.016). Given that BMI has already in part corrected for height, this may partially explain these values. BMI produced slightly higher R^2 values in comparison to height (0.032 to 0.041). Insulin produced the smallest global R^2 out of the variables and likely reflects that it explained the least amount of the total variation in its mPCA model compared to the other variables that explained greater than 5%. The R^2 values of the variables that did not reach significance were zero in many of the models and reflect the variables that explained less than 5% but greater than 2% of the total variation in their mPCA models (Tables 19 and 20).

6.5.3 Sex

In both models, with outliers included and excluded, sex influenced the cheeks and orbits in a positive direction, whilst the midface and brow-ridge were negatively influenced. These findings confirm the findings of the mPCA model, with female faces presenting with rounder faces, more prominent eyes, smaller noses and less prominent brow-ridges (Figures 62 and 63).

6.5.4 Height

In the models that included both the sexes, the nasal bridge, pronasale and the chin were influenced in a positive direction, whilst the cheeks were influenced in a negative direction. Therefore, subjects that were taller have larger noses and either a more prominent chin and/or increased facial height (constrained due to the scaling procedures). When the sexes were assessed separately, similar facial differences were visualised with some subtle differences. Pronasale appears to be less prominent in the female subjects, although still increased with increasing height. The cheeks were more strongly influenced in the male subjects. When weight was included in the PLSR model rather than BMI, height also appeared to influence the prominence of the lips (Figures 62 and 63).

6.5.5 BMI /Weight

In all the models, the facial differences were consistent with regards to BMI. As the BMI increases, the cheeks are larger and the supra-orbital regions more prominent.

Conversely, the nose and lips are relatively retrusive as well as the infra-orbital region. The underside of the chin also appears to be negatively influenced by BMI, which suggests that the lower facial height is reduced as BMI increases, but that this was constrained by the scaling procedures used. The use of weight rather than BMI in the models did not appear to affect the influence on the facial features (Figures 62 and 63).

6.5.6 Fasting insulin

When the sexes were assessed together, increased levels of insulin influenced the prominence of the supra-orbital regions and buccal fat pads. The chin is negatively influenced, whilst the nose also appears smaller with increasing insulin levels. When the sexes were assessed separately, the global *p*-values were no longer significant at a 0.05 level. In females, the brow-ridge appears to increase in prominence with increasing insulin levels, whilst the corners of the mouth are also influenced in a positive direction. The infra-orbital regions, and to a lesser extent the chin, are influenced in a negative direction. In the male subjects, the chin is influenced the most, again in a negative direction. The buccal fat pads are influenced in a positive direction, and to a greater extent than in the female subjects. The brow-ridge is less effected than in the females. The supra-orbital region is influenced in a positive direction, but only on the right eye, suggesting that increasing insulin levels may impact the development of facial asymmetry, particularly in the orbital region. However, as the separate sex models did not reach global significance with regards to insulin and no model reached significance when Bonferroni correction was considered, these findings are suggested with caution (Figures 62 and 63).

6.5.7 Influence of removing variables from the model

Removing the variables that explained less than 2% of the variation in their respective mPCA models did not appear to influence the results with regards to *p*-values, R^2 values or facial shell differences for the remaining variables (Tables 16-17, Tables 19-20, Figure 64).

Table 16: Global p-values for each variable using an initial PLSR model that included the variables that explained greater than 2% of the total variation in the two-level mPCA models [10,000 permutations].

Outliers	Model	Sex	Height	BMI	Insulin	Cholesterol	Triglycerides	VLDL	LDL	Smoking 1st Tri	Smoking 2nd Tri	Alcohol 1st Tri
Outliers Included	Both Sexes	< 0.00001*	< 0.00001*	< 0.00001*	0.018 º	0.393	0.654	0.655^	0.784^	0.568	0.494	0.090
Outliers Excluded	Both Sexes	< 0.00001*	< 0.00001*	< 0.00001*	0.027 º	0.403	0.582	0.586	0.678	0.573	0.502	0.091
	Females Only	-	0.005 º	< 0.00001*	0.051±	0.919	0.458	0.452	0.819	0.323	0.233	0.336
	Males Only	-	< 0.00001*	< 0.00001*	0.145	0.222	0.743	0.757	0.753	0.133	0.501	0.355

* significant after Bonferroni correction (0.05/11 = 0.0045)

º significant at 0.05 level

± almost significant at 0.05 level

^ VLDL and LDL were strongly correlated with cholesterol and triglycerides respectively when assessed using Pearson correlation. The model was therefore re-run without these variables as a check and the results remained the same.

Table 17: Global P values for each variable using a PLSR model that included the variables that explained greater than 5% of the total variation in the two-level mPCA models and were significant in the initial PLSR model. Their interactions were included initially and removed from subsequent models if found to be non-significant (those that were close to significant were kept initially).

Outliers	Number of Permutations Used	Model	Sex	Height	BMI	Insulin	Sex*Height	Sex*BMI	Sex*Insulin	Height*BMI	Height*Insulin	BMI*Insulin
Outliers Included	1,000 permutations	Both sexes	< 0.001*	< 0.001*	< 0.001*	0.013 º	0.081±	0.214	0.1±	0.222	0.163	0.126
	10,000 permutations		< 0.00001*	< 0.00001*	< 0.00001*	0.024 º	0.081	-	0.082	-	-	-
			< 0.00001*	< 0.00001**	< 0.00001*	0.016 º	-	-	-	-	-	-
Outliers Excluded	1,000 permutations	Both sexes	< 0.001*	< 0.001*	< 0.001*	0.046 º	0.06±	0.055±	0.158	0.189	0.148	0.201
	10,000 permutations	Both Sexes	< 0.00001*	< 0.00001*	< 0.00001*	0.035 º	0.093	0.015*	-	-	-	-
			< 0.00001*	0.00001*	< 0.00001*	0.036 º	-	0.020 º	-	-	-	-
			< 0.00001*	< 0.00001*	< 0.00001*	0.030 º	-	-	-	-	-	-
		Females Only	-	0.004*	< 0.00001*	0.102	-	-	-	-	-	-
		Males Only	-	< 0.00001*	< 0.00001*	0.096	-	-	-	-	-	-

* significant after Bonferroni correction (0.05/4 = 0.0125)

º significant at 0.05 level

± included to be conservative in subsequent model as almost significant at 0.05 level

Grey: model explored further via visualisations of the coefficients and R^2 facial meshes

Table 18: Global *p*-values for PLSR model assessing weight rather than BMI [10,000 permutations]

Outliers	Sex	Height	Weight	Insulin	Sex *Height	Sex *Weight	Sex *Insulin	Height *Weight	Height *Insulin	Weight *Insulin
Outliers Included	< 0.0001*	< 0.0001*	< 0.0001*	0.017 º	0.034 º	0.577	0.117	0.565	0.324	0.175
Outliers Excluded	< 0.0001*	< 0.0001*	< 0.0001*	0.049 º	0.017 º	0.286	0.135	0.360	0.294	0.173
	< 0.0001*	< 0.0001*	< 0.0001*	0.036 º	0.110	-	-	-	-	-
	< 0.0001*	< 0.0001*	< 0.0001*	0.034 º	-	-	-	-	-	-

* significant after Bonferroni correction (0.05/4 = 0.0125)

º significant at 0.05 level

Grey: model explored further via visualisations of the coefficients and R^2 facial meshes

Table 19: Global R^2 for each variable using an initial PLSR model that included the variables that explained greater than 2% of the total variation in the two-level mPCA models [10,000 permutations]

Outliers	Sex	Sex	Height	BMI	Insulin	Cholesterol	Triglycerides	VLDL	LDL	Smoking 1st Tri	Smoking 2nd Tri	Alcohol 1st Tri
Included	Both	0.043	0.008	0.037	0.002	0	0	0	0	0	0	0.001
Excluded	Both	0.043	0.008	0.034	0.002	0	0	0	0	0	0	0
	Females Only	-	0.004	0.032	0.002	0	0.001	0.001	0	0.001	0.002	0.001
	Males Only	-	0.016	0.038	0.002	0.002	0.001	0.001	0.001	0.002	0.001	0.002

Table 20: Global R^2 for each variable using a PLSR model that included the variables that explained greater than 5% of the total variation in the two-level mPCA models and were significant in the initial PLSR model [10,000 permutations]

Outliers	Sex	Sex	Height	BMI	Insulin	Sex*Height	Sex*BMI	Sex*Insulin	Height*BMI	Height*Insulin	BMI*Insulin
Included	Both	0.047	0.007	0.041	0.002	0.001	-	0.001	-	-	-
		0.047	0.009	0.041	0.002	-	-	-	-	-	-
Excluded	Both	0.045	0.008	0.037	0.001	0.001	0.002	-	-	-	-
		0.045	0.009	0.037	0.001	-	0.001	-	-	-	-
		0.047	0.009	0.037	0.002	-	-	-	-	-	-
	Females Only	-	0.004	0.037	0.002	-	-	-	-	-	-
	Males Only	-	0.018	0.039	0.003	-	-	-	-	-	-

Grey: model explored further via visualisations of the coefficients and R^2 facial meshes

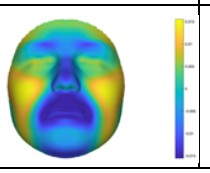
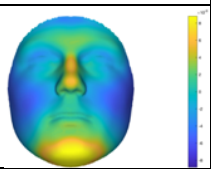
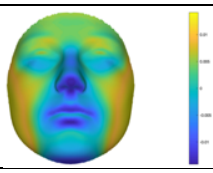
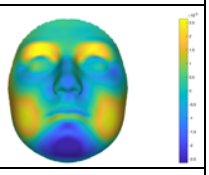
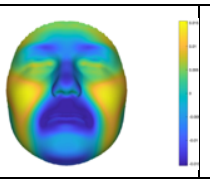
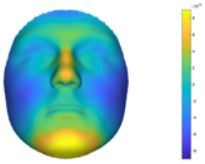
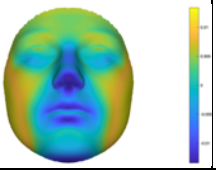
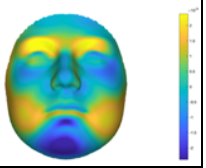
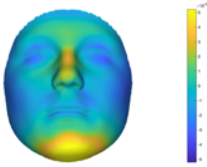
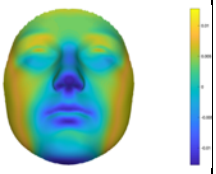
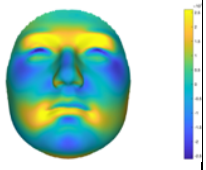
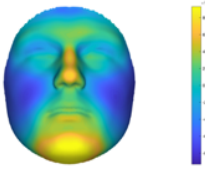
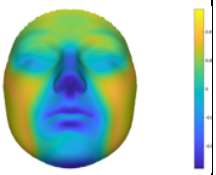
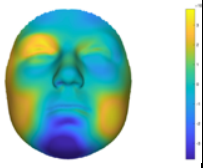
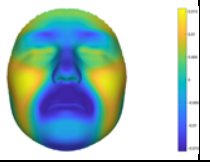
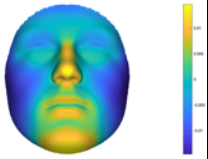
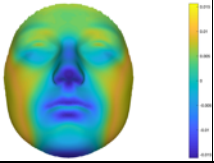
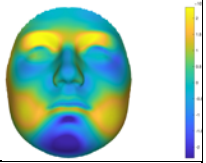
Outliers	Model	Sex	Height	BMI	Insulin
Included	Both Sexes				
					NB: global *p*-value not significant after Bonferroni correction
Excluded	Both sexes (Sex * BMI included in model)				
					NB: global *p*-value not significant after Bonferroni correction
	Females Only	-			
					NB: global *p*-value not significant
	Males Only	-			
					NB: global *p*-value not significant
		Sex	Height	Weight	Insulin
	Both sexes				
					NB: global *p*-value not significant after Bonferroni correction

Figure 62: PLSR model coefficients. Yellow = larger/ more prominent; blue = smaller, more retrusive.

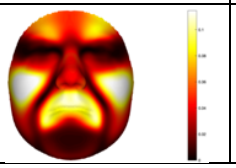
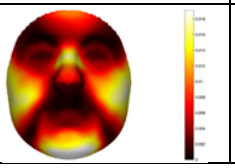
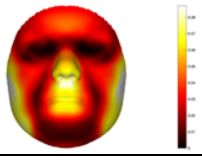
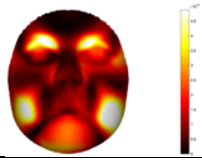
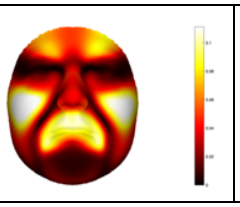
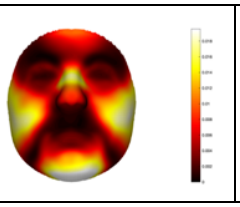
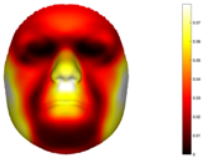
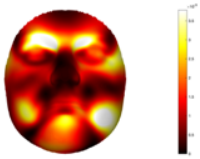
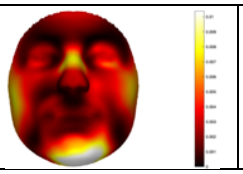
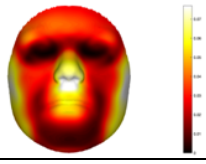
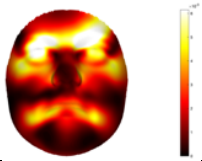
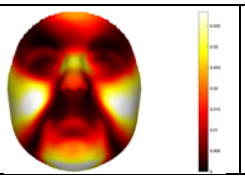
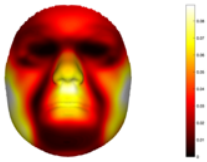
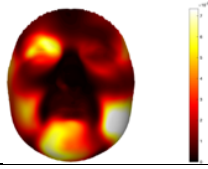
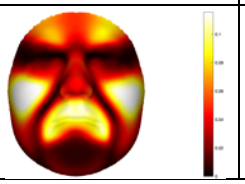
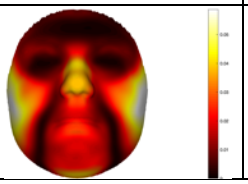
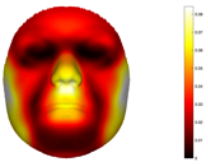
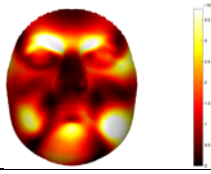
Outliers	Model	Sex	Height	BMI	Insulin
Included	Both sexes				
					NB: global *p*-value not significant after Bonferroni correction
Excluded	Both sexes (Sex * BMI interaction included in model)				
					NB: global *p*-value not significant after Bonferroni correction
	Females Only	-			
					NB: global *p*-value not significant
	Males Only	-			
					NB: global *p*-value not significant
		Sex	Height	Weight	Insulin
	Both sexes				
					NB: global *p*-value not significant after Bonferroni correction

Figure 63: Partial R^2 for each landmark in the PLSR models. White = higher partial R^2; black = lower partial R^2

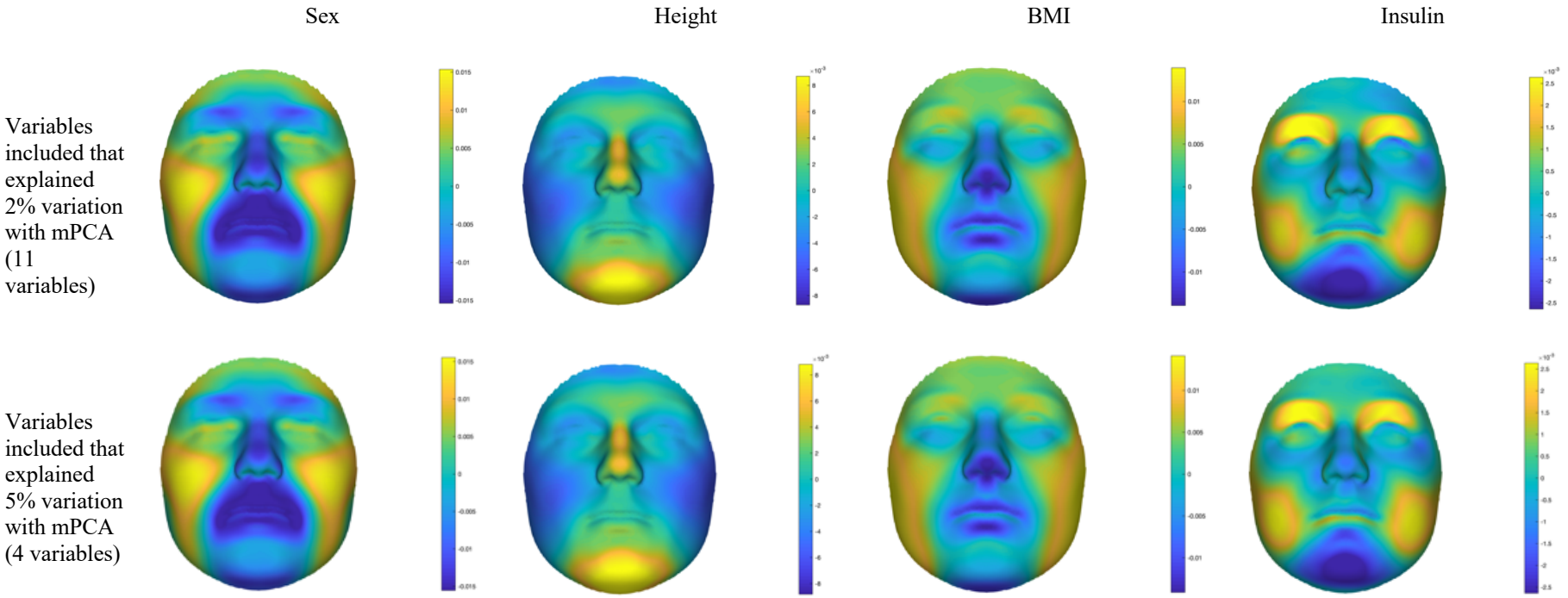


Figure 64: Comparing the facial shells (coefficients only) with variables reaching 2% variation explained in their mPCA model (11 variables) and 5% variation (4 variables) – representative faces used as multiple models explored.

6.5.8 Other (non-significant) variables

The findings of the variables without a significant global *p*-value are described for completeness and should be interpreted with caution. The visualisations are available in Figure 65. Increasing cholesterol appears to have a positive influence on the shape of the cheeks as well as the temporal region. The nose, upper lip and chin appear to be influenced in a negative direction. The influence on the supra-orbital region is difficult to interpret with certainty as different results are obtained when VLDL and LDL were removed from the model. The supra-orbital regions appear to be influenced in positive direction alongside the bridge of the nose. When VLDL and LDL were removed from the analysis, there was evidence of facial symmetry with the right of the mandible and left of the forehead influenced in a positive direction. The cheeks may be influenced in a negative direction, but this differs between the models. The cheeks and supra-orbital regions are influenced in a positive direction, whilst the bridge of the nose and mental fold are influenced in a negative direction. The results for LDL were different between the PLSR and mPCA models. In PLSR, the cheeks were influenced in a negative direction. In both models, the supra-orbital regions were influenced in a positive direction.

The results of the PLSR models with regards to maternal smoking are contradictory. Maternal smoking during the 1st trimester influenced the cheeks and upper lip in a positive direction, whilst the nose, forehead and chin were influenced in a negative direction. Conversely, the PLSR model suggests that maternal smoking during the 2nd trimester influences the forehead, bridge of nose and chin in a positive direction, whilst the cheeks and infra-orbital regions were influenced in a negative direction. Finally, maternal alcohol consumption during the 1st trimester influenced the cheeks in a positive direction. The supra-orbital regions were also influenced in this direction. The chin, lips and nose were influenced in a negative direction in the PLSR models.

6.5.9 Comparison of results: PLSR v mPCA

The variables that influenced facial shape the most (sex, height, BMI and fasting insulin) showed similarity in the results between the PLSR and mPCA models. However, differences between the results became more apparent for the variables that had less of an influence on facial shape (Table 21).

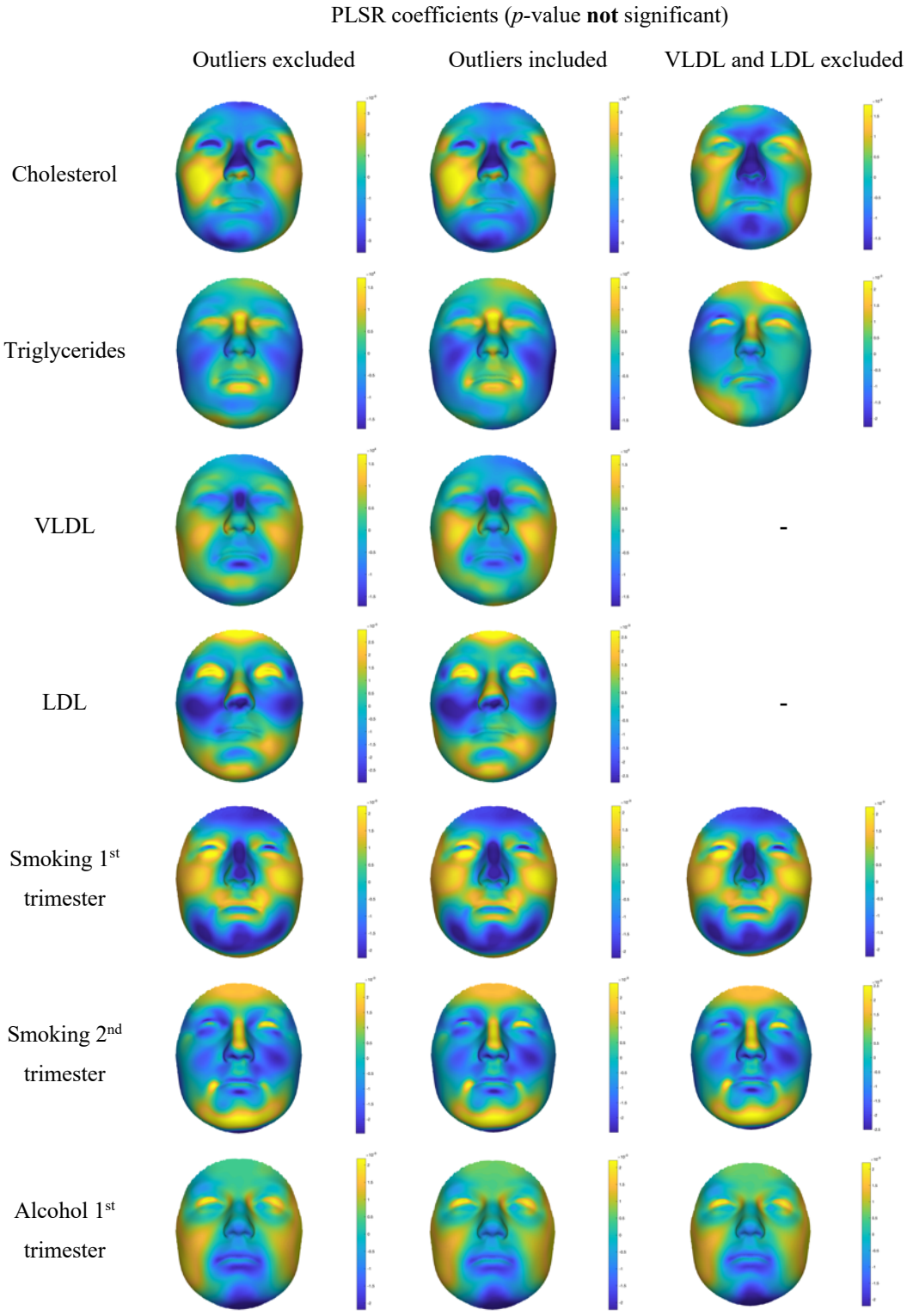
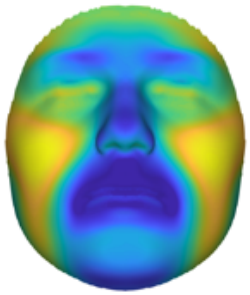
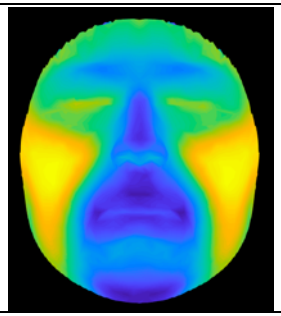
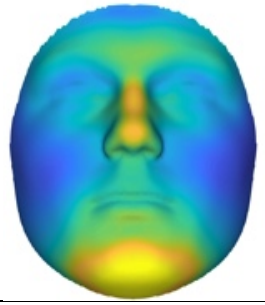
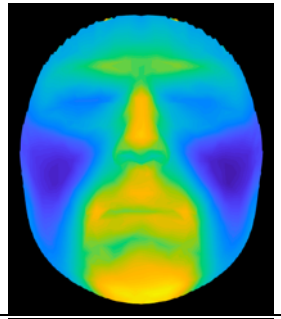
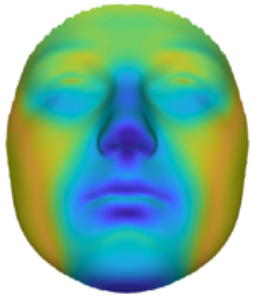
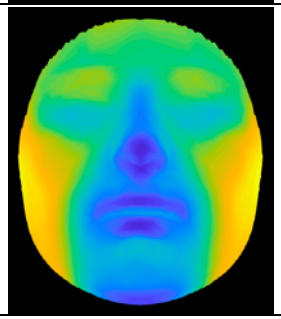
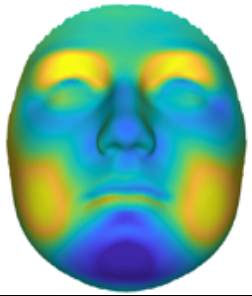
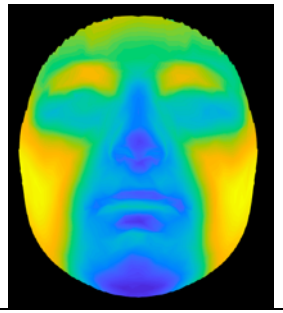
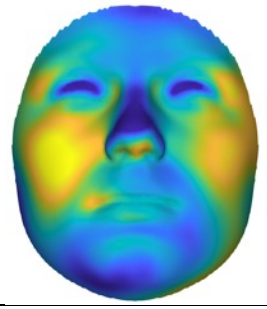
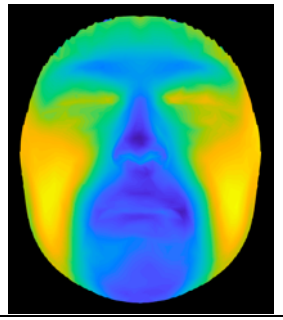
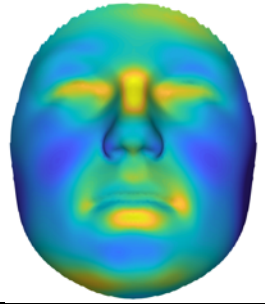
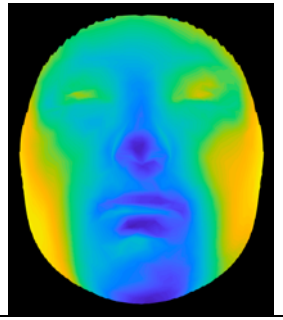
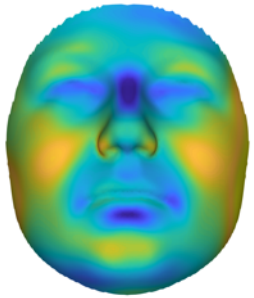
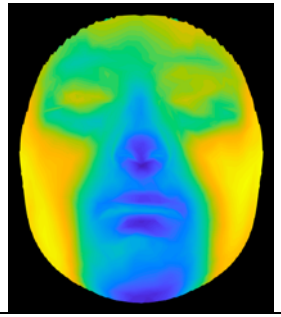
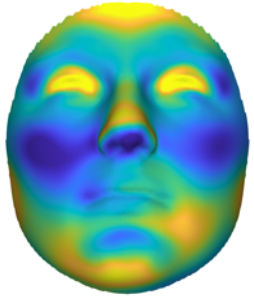
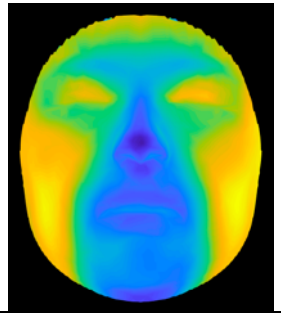
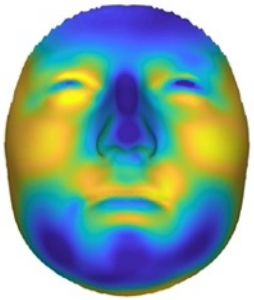
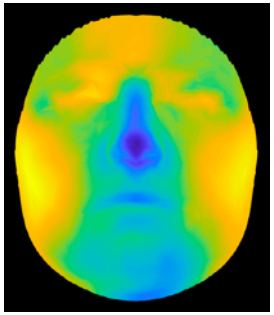


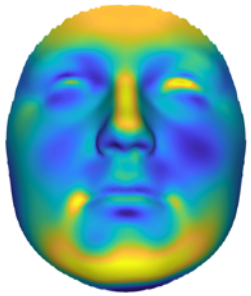
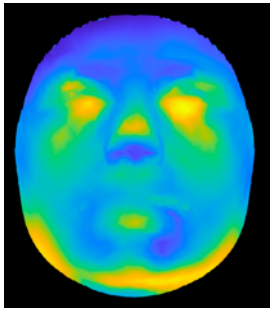
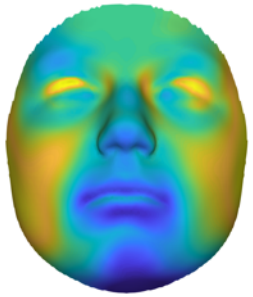
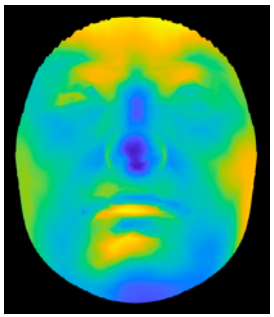
Figure 65: Coefficients of the variables that did not reach global significance in the PLSR models. The results of the mPCA models are presented for comparison. Yellow = larger/ more prominent; blue = smaller, more retrusive.

Table 21: Comparison of results from PLSR and mPCA (representative face chosen from PLSR models)

Variable	PLSR	mPCA (PC1)	Similarities	Differences
Sex			Cheeks (larger) Midface (retrusive) Infra-orbital regions (more prominent)	Minimal
Height			Nose (larger) Chin (larger) Brow ridge (larger) Cheeks (less prominent) Face narrower	Lips (more prominent) - mPCA
BMI			Cheeks (larger) Supra-orbital region (more prominent) Midface (retrusive) Forehead (more prominent)	Minimal

Fasting insulin			Cheeks (larger) Supra-orbital region (more prominent) Infra-orbital region (less prominent) Chin (retrusive) Nose (smaller)	Lower lip (more prominent) - PLSR Forehead (more prominent) - mPCA
Cholesterol			Cheeks (larger) Temporal region (more prominent) Nose (more retrusive)	Subtle differences in areas of cheeks affected Orbital regions (less prominent) - PLSR and Orbital regions (more prominent) - mPCA Lips (more retrusive) - mPCA
Triglycerides			Supra-orbital region (more prominent)	Disagreement in most features

VLDL			Cheeks (larger) Lips (more retrusive) Nasal bridge (more retrusive)	Forehead and chin (more prominent) - mPCA Supra-orbital region (more prominent) - mPCA
LDL			Supra-orbital regions (more prominent) Upper forehead (more prominent)	Disagreement in all other areas
Smoking 1st trimester			Nose (more retrusive) Outer area of cheeks (larger)	Disagreement in all other areas

Smoking 2nd trimester			Areas of mandible (more prominent) Lower part of bridge of nose (more prominent)	Disagreement in all other areas
Alcohol 1st trimester			Supra-orbital regions (more prominent) Cheeks (more prominent) Nose (smaller) Chin (less prominent)	Forehead (more prominent) – mPCA Lips (less prominent) – PLSR Lips (more prominent) - mPCA

6.6 DISCUSSION

As in studies one and two, discussion of the influence of each variable is provided in the General Discussion (7.1 Facial differences, p. 165). The ability to add multiple variables to the PLSR models, including both continuous and categorical variables, gave more flexibility than mPCA. The adjustment for variables within the PLSR model also made the results more reliable compared to independent mPCA models.

Sex, height, BMI and fasting insulin explained more than 5% of the total variation in their respective mPCA models and reached significance in the PLSR models (although fasting insulin became non-significant after Bonferroni correction). Those that explained less than 5% of the total variation in their respective mPCA models did not reach significance in the PLSR models. The *p*-values, R^2 values and facial differences with regards to sex, height, BMI and fasting insulin did not appear to be affected when these variables were included or excluded. The inclusion of these variables may therefore induce overfitting and increase computational time without any benefit.

6.7 SUMMARY

A summary of the inferential results from this study are presented in Table 22. Due to the lack of adjustment for variables in the mPCA models, the results from the PLSR models may be more reliable. However, the use of mPCA as a variable selection tool prior to PLSR was beneficial as the number of variables included in the model could be reduced significantly and thus reduced computational time as well as the risk of overfitting. It would appear sensible to suggest 5% variation explained in a mPCA model as a threshold with mPCA prior to inclusion PLSR models in future given these variables reached significance in the PLSR models.

Table 22: Summary of results from study three using PLSR with bootstrapping

Variable	*P*-value	Smallest threshold for significance from the models	Null hypothesis
Sex	<0.00001	0.0045	Reject
Height	<0.00001	0.0045	Reject
BMI	<0.00001	0.0045	Reject
Cholesterol	>0.05	0.0045	Accept
Triglycerides	>0.05	0.0045	Accept
VLDL	>0.05	0.0045	Accept
LDL	>0.05	0.0045	Accept
Insulin	0.017 – 0.049	0.0045	Accept
Smoking 1st trimester	>0.05	0.0045	Accept
Smoking 2nd trimester	>0.05	0.0045	Accept
Alcohol 1st trimester	>0.05	0.0045	Accept

7 GENERAL DISCUSSION

7.1 FACIAL DIFFERENCES

7.1.1 Overall summary of results

A summary of the total variation explained and inferential findings for each of the variables are provided in Table 23. The results for geographical location were based on 21 landmarks only whilst the other results are provided for 1000 quasi-landmarks (mPCA) and 7160 quasi-landmarks (PLSR).

Table 23: Summary of total variation and inferential tests for each variable

Variable	% Total variation explained (mPCA model)	Null hypothesis (mPCA)	Null hypothesis (PLSR)
Geographical location (21 landmarks)	11.34 – 14.59%	Reject	-
Sex	17.31%	Reject	Reject
Age	1.18%	Accept	-
Height	11.05%	Reject	Reject
BMI	17.29%	Reject	Reject
Pubertal stage (hair)	1.76%	Reject	-
Pubertal stage (genitals)	0.58%	Reject	-
Cholesterol	3.55%	Reject	Accept
Triglycerides	2.25%	Reject	Accept
VLDL	3.54%	Reject	Accept
LDL	2.01%	Reject	Accept
HDL	1.58%	Reject	-
Glucose	1.69%	Reject	-
Insulin	5.83%	Reject	Accept (after Bonferroni)
Asthma (0-42 months)	0.93%	Reject	-
Asthma (91 months)	0.40%	Reject	-
Atopy	0.15%	Reject	-
Hay fever	0.35%	Reject	-
Smoking before pregnancy	1.53%	Reject	-
Smoking 1st trimester	2.51%	Reject	Accept
Smoking 2nd trimester	3.74%	Reject	Accept
Alcohol before pregnancy	1.56%	Reject	-
Alcohol 1st trimester	2.62%	Reject	Accept
Alcohol 2nd trimester	1.60%	Reject	-

7.1.2 Geographical location

It appears that the geographical location of a subject explains more of the differences in the faces than the sex of the subject (both when the Croatian sample, who were older, were included and excluded) for the subject groups studied here. This contrasts with the thoughts of Scheuer (2002) who advised that ethnicity is the more difficult to ascertain than sex from the skeleton. This may be due to differences in analysing living individuals via facial scans and dry skulls. However, the differences at this level may be exaggerated in this thesis due to the different researchers placing the landmarks on each sample. Although it is hoped this issue was reduced via calibration. The influence of sex may also be underestimated here due to the restriction in the number of eigenvalues possible due to there being only two groups present (female and male).

Differences due to geographical location can be seen in most of the main facial features. The facial features that have been found to differ in different European populations also appear to be influenced in this study. Previous research suggests that forehead shape, nose length and width, width of eyes, lip prominence, mandibular shape and chin prominence all differ between populations (Bozic et al. 2009; Kau et al. 2010; Hopman et al. 2014; Farnell et al. 2017). Differences in nose shape, eye width and mouth shape were highlighted in particular in this thesis.

As all the populations in this study are Caucasian, it is encouraging that differences between the groups were found. The Croatian sample appears to be most distinguishable. This is likely due to their older age than the other populations. It is not surprising that the multilevel PCs showed no substantial areas of variation between the English (Bristol) and Welsh (Cardiff) populations given their proximity geographically. They are therefore likely to have been exposed to similar environmental factors and possibly genetics. Indeed, a recent GWAS study suggests that some populations in Cardiff area (although not all) may have similar genetics to those in the Avon area (Leslie et al. 2015). There was clear clustering separating the centroids of the two populations, particularly for mPCA population level PC2 and PC3. This indicates that subtle differences were found.

Overall, the findings support Hopman et al. (2014), who suggest that combining multiple populations in studies into a broad racial category, such as Caucasian, may be inappropriate due to the differences between subpopulations. Hopman et al. (2014) also suggest that there is likely to be a high level of genetic admixture in populations that have undergone immigration of other populations throughout their history. This holds true for the populations investigated here, and perhaps reflects the large amount of within-group variation seen.

7.1.3 Sex

Following puberty, the skull is the most sexually dimorphic region of the body after the pelvis (Scheuer and Black 2007b). It follows that the facial features also appear to be highly sexually dimorphic with many studies being able to differentiate between the faces of males and females to a high level of sensitivity (for example, (Abbas et al. 2018)). In the literature, female faces are suggested to be smaller and rounder, with less prominent brow-ridges, more prominent and wider apart orbits, more prominent cheeks, smaller noses, and fuller lips, (for example, (O'Toole et al. 1997; Graw et al. 1999; Nute and Moss 2000; Kau et al. 2006; Bozic et al. 2009; Gor et al. 2010; Velemínská et al. 2012; Bugaighis et al. 2013; Koudelová et al. 2015; Mydlová et al. 2015)). The findings in both the mPCA and PLSR models agree with all the findings, which is encouraging as it shows that mPCA produces meaningful results. The agreement amongst all the analyses provides further validity.

7.1.4 Age and pubertal stage

This thesis was not focussed on assessing growth or the influence of age long term. The subjects were all between 14-16 years old (except for the Croatian population in study one, who were included to assess the ability of mPCA to separate groups). Age was therefore assessed as a potential confounder only. As the age range was small, it was expected that the variation explained would be minimal. This could provide a further use for mPCA; assessing whether confounding factors have been adequately controlled in the sampling phase.

The mPCA model suggests that a reduced pubertal stage was associated with a retrusive chin or reduced lower facial height, less prominent forehead and more prominent cheeks, eyes, and lips. As these facial features are likely to present in this manner prior to puberty, particularly in males (Koudelová et al. 2015), it is important that studies investigating age also consider subjects' pubertal stage in future.

7.1.5 Height

Both the mPCA and PLSR models suggest that taller subjects present with more prominent brow-ridges, lips, and larger noses, with relatively narrower and potentially longer faces. Previous studies investigating the influence of growth hormone suggest that those with increased growth hormone present with an increased facial height. However, they also suggest that increased growth hormone is associated with increased facial width (Pirinen et al. 1994). It would therefore be worthwhile assessing the influence of growth hormone on the facial shape of the ALSPAC cohort. However, this is not available for the subjects at the time of their facial scans.

7.1.6 BMI

Previous studies suggest that an increased BMI is associated with increased facial width and increased mandibular width and length (Ohrn et al. 2002; Ferrario et al. 2004; Sadeghianrizi et al. 2005). There is however controversy in the literature as to whether obesity is linked to a retrusive or protrusive profile and whether the anterior facial height is increased or reduced (Ohrn et al. 2002; Ferrario et al. 2004). Here, the mPCA and PLSR models agree that obesity is associated with larger cheeks and therefore increased facial width. The midface appears to be retrusive in comparison. However, it is challenging to determine whether this is due to a true clinical retrusive midface or a relative retrusion in comparison to the size of the cheeks given that the faces are aligned using GPA without a specific, stable reference point. The models suggest that the anterior facial height is reduced. However, further investigation is required with regards to the *x*-, *y*- and *z*- axes individually. The findings at a BMI level make sense, in that adipose deposits are expected at the buccal fat pads. Adipose tissue deposition can also be found in the orbits (Wolfram-Gabel and Kahn 2002) and may in part explain the influence of BMI on orbital shape.

7.1.7 Fasting insulin

The metabolic factor that appears to be most promising for further investigation is fasting insulin, as differences were described in both the mPCA and PLSR models. Although the results did not reach significance in the PLSR models, which in part may be due to the conservative nature of Bonferroni correction. A previous study suggested that insulin affects facial height, nasal asymmetry, and the depth of the upper eyelids (Djordjevic et al. 2013b). Here, the mPCA model suggests that insulin has an influence on many more of the facial features. High levels of insulin appear to influence the cheeks, forehead and supra-orbital regions in a positive direction, and the nose, infra-orbital regions, lips, chin and/or lower facial height in a negative direction. The influence on female and male subjects appears to differ in the PLSR models.

There is biological reasoning behind the influence of insulin on facial shape, given the complex relationship between insulin and GH suggested by Qiu et al. (2017). The positive influence on the brow-ridge may be explained if insulin optimally stimulates GH. Qiu et al. (2017) suggests this would be expected in non-obese and non-diabetic individuals. Growth at the cranial base can in part be compared to growth in the long bones due to the similarities between the synostoses in the cranial base and the epiphyses in the long bones. They may therefore be sensitive to the same factors. However, the pre-sphenoid and cribriform plate regions are considered stable from seven years old with the movement of nasion in forwards trajectory after this age associated with an increase in size of the frontal sinus (Afrand et al. 2014). It follows that any influence of insulin on the anterior cranial base is likely to have occurred prior to the age of seven. Further work is therefore required to assess whether the potential difference in brow-ridge prominence occurred prior to the age of seven, and was maintained until 15 years old, or whether these differences appeared after seven years old and are therefore more attributable to expansion of the frontal sinus.

The negative influence on the mandible and subsequent class II skeletal pattern could be attributed to an increased length of the anterior and/or posterior cranial base (Hopkin et al. 1968; Dibbets 1996). However, the association between cranial base length and skeletal pattern is conflicting in the literature, with other studies finding no

significant relationship (Kasai et al. 1995; Wilhelm et al. 2001; Polat and Kaya 2007; Kamak et al. 2013; Breh and Kamat 2015). It is more challenging to explain a positive influence on growth at the brow-ridge and a negative influence on mandibular prominence without implicating the cranial base. However, it is interesting that males appear to be more sensitive to insulin at the mandible (negative association) and females at the brow-ridge (positive association), albeit at levels below statistical significance. Further work is required to investigate this further.

It will be challenging to attempt to disentangle the influence of sex, obesity, type II diabetes, insulin levels, growth hormone levels and height in future models. Given the constraints on the number of levels possible in mPCA due to sample sizes, it is unclear whether the differences seen here are due to insulin or whether they are due to confounding factors. However, given that differences were also seen in the PLSR model, albeit in fewer facial features, there is evidence to suggest that insulin is worth exploring further. Exploring the association between genetic factors, insulin and facial shape may also be of interest. SNPs close to the following have been associated with insulin resistance. However, it does not appear that they are associated with any facial feature at present: Chromosome17p, *VNTR, PPARG, KLF14, IRS1, GCKR, FTO, TCF7L2, NAT2, TMEM163, IGF1, MC4R, SC4MOL, TCERG1L* and *ARL15* (Le Stunff et al. 2000; Rich et al. 2005; Brown and Walker 2016).

It is interesting that the PLSR results for insulin were significant for all analyses (prior to Bonferroni correction) when both sexes were combined but were not significant when females and males were separated. This phenomenon may be linked to Simpson's Paradox (Simpson 1951). Simpson's Paradox describes a situation where a result is significant when assessing subgroups, but when the groups are assessed together, the result becomes non-significant. Here, the opposite appears to be true, as analysing the sexes together resulted in a significant result, but when the sexes were assessed separately, the results were not significant. A visualisation is proposed in Figure 66. However, this requires further investigation.

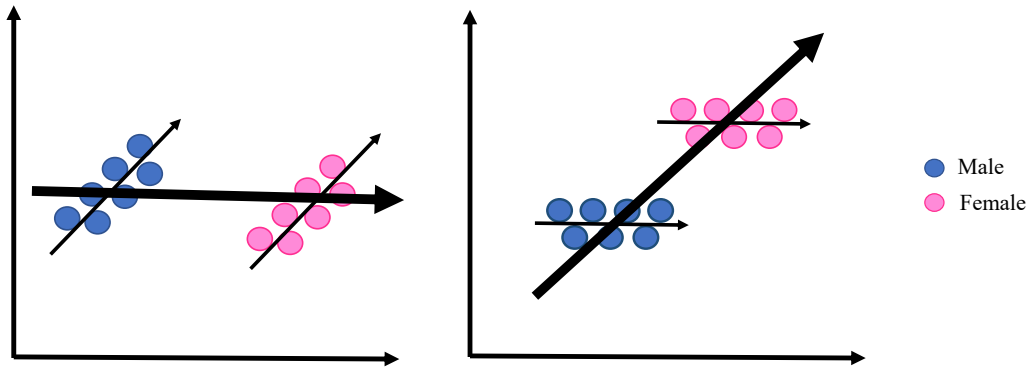


Figure 66: (a) Simpson's paradox: when the groups are analysed separately, there is a positive relationship between *x* and *y*. However, when the groups are analysed together, there is no correlation. (b) A visualisation exaggerating what might be happening with regards to insulin in the PLSR models: when the groups are analysed separately, there is no relationship between x and y. However, when there are analysed together, there is a positive relationship.

7.1.8 Glucose, cholesterol, VLDL and HDL

There does not appear to be any published findings on the influence of glucose, cholesterol, VLDL or HDL, apart from Djordjevic et al. (2013b), who suggested that these variables were not associated with any of the PCs in their study at a statistically significant level. Further work, using different models other than mPCA may therefore be useful in future. Djordjevic et al. (2013b) did however suggest that triglycerides were associated with PCs that explained nose height. The findings here, in the mPCA model, suggest that the cheeks are larger as well as the supra-orbital region and forehead when triglyceride levels are increased. In contrast, the nose was smaller alongside the lips and chin. However, as the findings do not agree with the PLSR model, and the findings in the PLSR model did not reach significance, further work is required to verify these findings, particularly as confounding factors were not corrected for in the mPCA model. LDL was also investigated by Djordjevic et al. (2013b), who found that LDL was associated with PCs explaining nasal prominence, lower lip prominence and nasal asymmetry. Here, most of the facial features were affected and again further work is needed to verify these findings.

7.1.9 Breathing disorders

Of all the variables assessed in this thesis, breathing disorders appear to have the smallest influence on facial shape. However, given the intimate relationship between

the respiratory system and the face, it would be surprising if there was no association. Al Ali et al. (2014b) were able to use a larger sample from the ALSPAC cohort as they did not explore as many variables and therefore did not need to exclude as many subjects due to missing data. They found significant differences in the facial shape of subjects whose mothers reported that they wheezed at 7.5 years old. These differences were in nose width, mid-face angle and a reduced height of the mid-face in females only. However, these differences were in the region of 0.4-0.9mm. These findings are like those found here, with the mPCA model suggesting that asthma at 7.5 years old resulted in a more retrusive and/or reduced mid-facial height. In addition, the orbits and mandible appear to be influenced as well. Wheezing between 0-3.5 years old was also assessed here, with subjects presenting with smaller noses. In addition, there was a suggestion that the lower facial height may be increased, which would agree with the suggestions an elongated face could be associated with a reliance on mouth breathing rather than nasal breathing (Linder-Aronson 1974)

Atopy explained the least amount of total variation in this thesis. It was therefore not explored further in the PLSR models. In the mPCA model, atopy appeared to be associated with retrusive or reduced midfacial proportions, larger noses, and foreheads, including the brow-ridge. The mandible also appeared to be affected and well as facial symmetry. However, as atopy influenced the total variation the least, these results should be interpreted with caution. A previous study on the ALSPAC cohort suggests that facial heights were longer in subjects with atopy (0.4-0.6mm) (Al Ali et al. 2014a). Again, a larger sample size was possible as fewer variables were assessed. Assessment of unscaled faces would therefore be useful in an mPCA model to determine the influence of atopy on facial height.

A historical study investigating the influence of perennial allergic rhinitis on the facial shape of 5 to 10 year olds suggests that the Frankfort Mandibular Planes Angle and lower facial height are increased whilst the face becomes narrower. They suggested that further work was required to assess whether treatment influences these facial features (Sassouni et al. 1985). Trask et al. (1987) also assessed the influence of perennial allergic rhinitis by comparing patients with their non-allergic siblings and controls who were nasal breathers. They suggested that hay fever results in longer facial shapes and that the allergic patients as well as their siblings presented with

retrusive facial profiles. The results of the mPCA model agree that the maxillary region may be retrusive in nature, with a relatively more prominent pronasale. Further work disentangling the influence of *x*-, *y*- and *z*- axes in the mPCA model would help determine if the differences are due to retrusive profiles or whether the mid-facial height is reduced or both.

Other breathing disorders that would be worth exploring in future, which were not explored here include, the influence of snoring, sleep apnoea and mouth breathing were not assessed here. These have been found to result in statistically significant increased face height, decreased nasal prominence and width, and mandibular retrognathia, in a study of the ALSPAC 3D facial scans. However, these differences were small, in the region of 0.3mm (Al Ali et al. 2015). Given that sleep apnoea is associated with obesity (Allison and Saunders 2000), insulin resistance, and dyslipidemia (Romero-Corral et al. 2010), investigating this may be worthwhile in future.

7.1.10 Maternal smoking during pregnancy

In this population, it appears that maternal smoking before pregnancy, during the first trimester and during the second trimester had a small effect on facial shape of the main facial features of the offspring at 15 years old. This effect was significant in the mPCA models, but the influence of smoking in the 1st and 2nd trimesters did not reach significance in the PLSR model. The mPCA models suggest that the nose and maxilla are smaller when mothers smoke during pregnancy in contrast to the cheeks. These facial features appear more pronounced if smoking continues through pregnancy. These results agree in part with the findings of Muggli et al. (2017), who suggested that the nasal bridge was less prominent in one year old, as well as the forehead and lower lip. They also found that the chin was more superior. If the facial shapes of the ALSPAC cohort had been assessed from birth, variations may have been more apparent during earlier stages of development. Certainly, it appears from the limited research available in the literature that the influence of maternal smoking reduces with age, particularly in females (Koziel 2018). This may be due to the influence of any exposure factors in utero being negated by the vast array of environmental factors they have been exposed to since birth.

7.1.10.1 DNA methylation

DNA methylation may be the mechanism through which maternal smoking influences foetal development, with *GFI1*, *KLF13*, *ATP9A*, *AHRR*, *MYO1G*, *CYP1A1* and *CNTNAP2* implicated in the ALSPAC cohort (Richmond et al. 2015). Although these genes are involved in development, they do not appear to have a specific influence on the development of facial structures, with the except of CYP1A1 and clefts (Shi et al. 2007). Thus, it may be expected that at a maternal smoking level, there is little difference in the main facial features of non-cleft patients. Maternal smoking has also been associated with a 33 SNPs in cleft patients (Beaty et al. 2013). It would be worthwhile reassessing whether these also have an influence on non-cleft facial features as GWAS research continues.

7.1.10.2 Smoking cessation

As would be expected, many of the mothers reported reducing their smoking or completely stopping smoking after learning of their pregnancy. 30.5% of mothers reporting stopping smoking before the first trimester and a further 18.5% before the 2nd trimester. This level of abstinence is possibly slightly higher than other studies, with an American paper stating that 22% of women stopped smoking after learning of pregnancy and 62% reduced their smoking levels (Heil et al., 2014). This difference could be attributed to the potential bias incurred due to the self-reported and retrospective nature of the maternal smoking status in this thesis. The results of the investigation into the cessation of smoking during pregnancy have not been presented in this thesis. This was explored as a four-level mPCA model with sex and maternal alcohol consumption in Galloway et al. (2020). It was not possible to assess the impact of smoking cessation and numbers of cigarettes concurrently due to sample sizes. The four-level mPCA model therefore focussed on any smoking or alcohol consumption before pregnancy, during the 1st trimester and during the 2nd trimester. The percentage explained by the maternal smoking level was minimal (0.66%). However, there was a suggestion from the component scores that the faces of the subjects whose mothers did not smoking during pregnancy were differentiable at PC2 from those who did smoke. Using alternative methodologies would be helpful for validating this finding. Furthermore, information is available from ALSPAC on the third trimester, which would also be interesting to explore. This is of particular interest as head

circumference was found to normalise if mothers stopped smoking by the third trimester (Lindley et al. 2000; Andersen et al. 2009).

7.1.10.3 Smoking topography

It may also be important to consider smoking topography as part of any future analysis as this may influence the levels of toxicity experienced by a foetus. Smoking tomography is an umbrella term describing how cigarettes are smoked in relation to the number of puffs that are taken of each cigarette; the length of each puff; the time between puffs; the volume of smoke consumed per puff; and the puff velocity (Bergeria et al. 2017). However, although it could be suggested that pregnant women may be more at risk of changing their smoking tomography due to abrupt smoking cessation and increased nicotine metabolism, no differences were found between pregnant and non-pregnant women (Bergeria et al. 2017). Furthermore, the influence of paternal smoking during development may have a minimal influence on foetal development as levels of plasma cotinine in non-smoking mothers with smoking partners in the ALSPAC cohort was low (Taylor et al. 2014). However, including these levels in any further analyses would give a thorough account of nicotine exposure in utero.

Additionally, the use of nicotine replacement therapy was not assessed in this study, which would expose the subjects to nicotine in utero, despite reduced maternal smoking levels. It would be interesting to investigate this in future, particularly as there is evidence of determinantal effects in animal studies (Bruin et al. 2010). However, a recent randomised trial on the use of nicotine-replacement therapy in an English population showed that compliance with nicotine-replacement therapy was low and did not significantly reduce the number of mothers who stopped smoking (Coleman et al. 2012). Nicotine-replacement may therefore have minimal impact on the ALSPAC sample. However, this should be taken into consideration to ensure robust results.

7.1.10.4 Smoking post-partum

Exposure to cigarette smoke after birth may also have an influence on development. It would also be useful to investigate maternal smoking levels during breastfeeding,

as nicotine is present in the breast milk of mothers who smoke and the production of breast milk is affected by smoking (Einarson and Riordan 2009). Furthermore, Stiby et al. (2013) found that in the ALSPAC population, plasma cotinine levels were significantly higher in 7 year olds and 15 year olds whose mothers are heavy smokers. Additionally, it has been found in an American study the children of smokers are more likely to smoke themselves (Vuolo and Staff 2013), even in a time of declining smoking. A study on the ALSPAC population has found that weekly smoking in adolescence can cause DNA methylation of *AHRR* (Prince et al. 2019). DNA methylation of *AHRR* was also found to be related to maternal smoking in a previous study (Richmond et al. 2018) thus highlights the importance of considering smoking in the adolescent as well as maternal smoking. The smoking levels of the children and exposure following pregnancy was not assessed in this study although may have a bearing on facial development, and thus would be useful to investigate in future.

7.1.10.5 Protective factors

Furthermore, protective factors, such as folic acid intake, may have reduced the influence of maternal smoking. The protective nature of folic acid was suggested by Carmichael et al. (2008) when assessing the influence of maternal smoking on craniosynostosis, although sample sizes were small and the confidence intervals all crossed one, thus suggesting a non-significant result. Multivitamins have also been suggested as a protective factor in the literature (Shaw et al. 2002). However it is difficult to assess this in isolation due to the large number of potentially confounding factors (Johnson and Little 2008), thus it would be interesting to assess potential protective factors as levels in the mPCA model in future studies.

7.1.11 Maternal alcohol consumption during pregnancy

Individuals with FAS present with a small nose, short palpebral fissures, a short, smooth upper lip and maxillary retrusion (Riley et al. 2011). Historically, FAS is associated with binge levels of alcohol consumption during pregnancy. Muggli et al. (2017) found that low levels of alcohol influenced forehead and nasal shape; moderate levels influenced orbital shape, the midface and chin shape; whilst binge levels influence chin shape. However, Howe (2019) found that the facial differences did

not reach significance using regression. The mPCA model suggests that alcohol consumption during pregnancy influences the nose (smaller), orbital shape and maxillary retrusion. This agrees with the clinical presentation of FAS and in part with the findings of Muggli et al. (2017). However, in addition to these facial features, the mPCA models also suggest that chin shape or lower facial height may be affected alongside forehead shape. There is also a suggestion of facial asymmetry in the subjects whose mothers consumed alcohol during the 1st and 2nd trimesters. Interestingly, maternal alcohol consumption before pregnancy also showed an influence on nose size and orbital shape, as well as forehead and chin shape. Determining whether this is due to epigenetic factors or due to alcohol consumption prior to knowledge of the pregnancy would be of interest in future work.

As was discussed for maternal smoking during pregnancy, the influence of stopping alcohol consumption during pregnancy was not explicitly explored here in a single model. This was investigated as a preliminary study using a four-level mPCA model with sex and maternal smoking. It was found that maternal alcohol consumption explained 0.48% of the variation, with little difference in the mean component scores of the groups (Galloway et al. 2020). Further investigation would be worthwhile with other methodologies to improve the reliability of this finding. Furthermore, many of the subjects reported drinking 1-2 glasses of alcohol per week. Further work on higher levels of alcohol consumption throughout pregnancy would be merited.

7.1.12 Within-group variation

The within-group variation level (all variation except that modelled in the other levels of the mPCA model) explained the greatest amount of variation in all of the models. This is expected due to the large variation in the faces we encounter daily. A recent GWAS has found in excess of 200 SNPs that are associated with facial shape (White et al. 2021) and it is likely that the number will continue to increase as GWAS research is replicated over time. This influence of the numerous genetic factors on facial shape are likely to explain a large proportion of the within group variation seen in this study. It is also likely that there will be numerous other environmental factors influencing that variation at this level, either in isolation or in association with genetic factors through epigenetics and shared genetics.

Although calibration was undertaken by the researchers involved in placing the 21 landmarks on the facial scans, plotting error could also contribute to variation between individual subjects within each group. In 3D laser facial scans, the orbit has been found to be one of the most difficult areas to landmark (Toma et al. 2009), whilst lasers have difficulty scanning the region due to the complex anatomy (Papadopoulos et al. 2002). These issues may also explain the high levels of within-group variation at the orbits and highlights the benefit of being able to separate this variation from the between group variation. Where quasi-landmarks were used (in analyses with 1000 landmarks and the PLSR analysis), there may be some error in the landmark positions due to the constraints of the methodology when warping the faces to fit the facial template. The ability to ascertain these differences separately to those associated with sex and population is essential in discerning subtle differences between faces, thus making mPCA an invaluable tool in 3D laser facial scanning research.

7.2 COMPARISON OF THE ANALYSES USED

This thesis used conventional PCA, DFA, PLSR and mPCA, with the aim of highlighting the relative advantages and disadvantages of each of these techniques in facial shape research. Figure 67 provides a visual summary of the techniques used in each of the three studies in this thesis. A comparison of the techniques can be found in Table 24.

Table 24: A comparison of the dimension reduction techniques used in this thesis

	Conventional PCA	DFA	PLSR	mPCA
Supervised?	No	Yes	Yes	Yes
Type of data (variables)	N/A	Categorical	PLSR: Continuous PLS-DA: Categorical	Categorical
Number of variables	Unlimited as not considered	Limited to one at a time	Unlimited but watch for overfitting	Usually 2-3 between-groups levels due to sample sizes
Sample size restrictions	More than one landmark	Number of subjects per group 4x number of retained PCs	Presumed more relaxed but debate in literature	No groups with 0 subjects, groups must have similar sample sizes
Distribution	Disagreement in literature	Normal distribution of landmarks	Normally distributed (although not essential)	Normal distribution of landmarks
Covariance matrix requirements	Nil	Homogeneity of covariance matrices required	Unclear	Homogeneity of covariance matrices to ensure sensible averaging at each level
Sensitivity to outliers	Sensitive	Sensitive	Very sensitive	Reduced with robust methods and use of median of covariance matrices
Descriptive statistics	Difficult to interpret	Very clear but difficult to interpret clinical differences	Difficult to interpret	Very clear
Inference via	MANOVA Multiple regression	Wilks Lambda	Bootstrapping partial R^2	MANOVA or ANOVA +/- bootstrapping

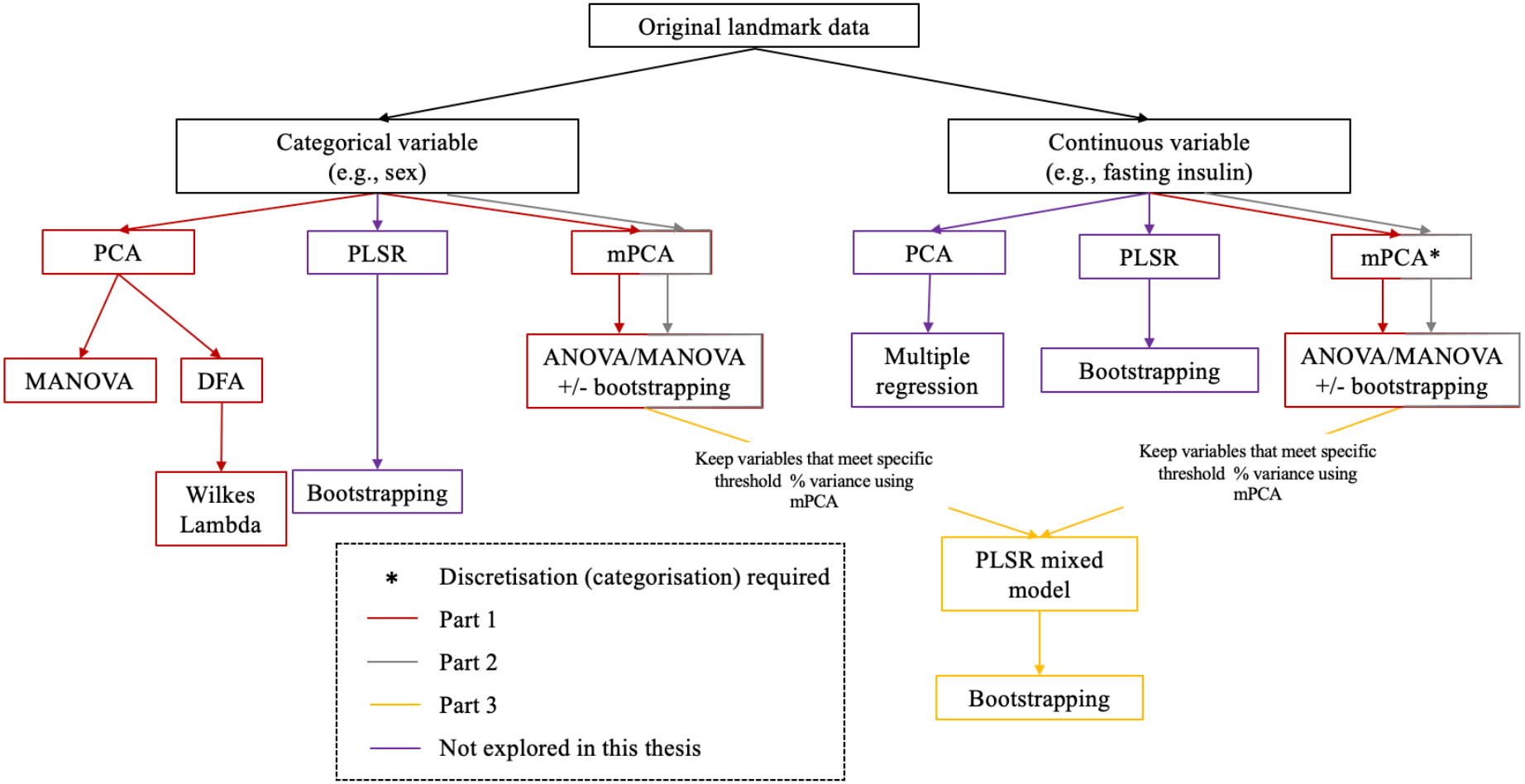


Figure 67: A flow chart describing the analyses used in this thesis.

7.2.1 Landmarks only

When no dimension reduction is carried out, detailed information on the differences in each of the landmarks is possible and there is an indication of both the between- and within-group variation. This can be simplified by visualising the mean of each landmark for each group with the addition of the standard deviation providing a summary of the within-group variation. However, it is difficult to visualise differences clearly due to the number of data points to process. It is slightly easier when only two groups are assessed but is challenging when the number of groups increases. Separate plots of each landmark would determine differences in more detail; however, these would be labour intensive to generate and analyse. Visualising the overall mean allows for a quick assessment of whether the populations are broadly different to each other. However, due to the processing stage, where the faces are superimposed using the mean in GPA, the mean of the landmarks no longer separates the groups. Given the caution required in the presence of multicollinearity and relatively small sample sizes, MANOVA was not used in these analyses. Until sample sizes are adequately increased with cohort studies encompassing very large populations, dimension reduction will be recommended in facial shape research.

7.2.2 Conventional Principal Component Analysis

The use of conventional PCA reduces the outcome variables into a more manageable number of PCs. Where the number of retained PCs is less than the number of subjects in each group, MANOVA can be used, with the hope that during the dimension reduction process little meaningful information is discarded. However, as conventional PCA is unsupervised, the variation explained by each PC is due to the mixing of multiple variables. It is therefore not possible to be certain that variation due to a variable of interest has not been discarded in part or in full during the dimension reduction process. The variation of interest is also mixed with within group variation. This is reflected in the similarity between the eigenvalue plots for conventional PCA and the within-group variation level of mPCA. These plots follow a similar pattern, indicating that much of the variation explained by each conventional PC is within group variation, therefore reducing their usefulness in visualising differences between the groups.

The number of score plots that require visualising increases in comparison to the raw landmarks. These score plots remain the same for each variable and are simply colour coded differently to visualise differences due to the different variables. This makes conventional PCA less useful from a data visualisation standpoint. It is also challenging, if not impossible, to attribute an overall influence on one variable as the results of each PC need to be combined in some way (whilst considering that the variation at each PC is attributable to multiple variables to greater and less extents). For example, if PC1 explains 18.57% of the total variation, PC2 explains 10%, PC3 explains 5% etc, these percentages cannot simply be added together to gain an overall importance for geographical location or sex. It is for these reasons that conventional PCA has significant disadvantages in facial shape research.

7.2.3 Discriminant Function Analysis

DFA is a method of separating the groups whilst providing a smaller number of plots for visualisation. As the process takes groupings into account, it is far easier than conventional PCA to determine the differences between groups. In its conventional form, DFA is only able to assess the influence of one variable at a time (although there is no restriction on the number of groups attributed to each variable). It would be interesting to assess DFA as a multilevel model. Furthermore, as DFA has sample size restrictions, conventional PCA was required as the number of landmarks often exceeded the necessary sample size for each group. The use of two separate dimension reduction techniques consecutively makes interpretation of the true facial differences more challenging than the use of conventional PCA in isolation. Finally, although 100% of the total variation explained by each DF is provided, this information is not useful for determining the overall influence of each variable. For example, if only one DF is possible (e.g., when determining the influence of sex), this DF explains 100% of the variation in the model. This information therefore has minimal clinical use.

7.2.4 Partial Least Squares Regression

The PLSR model used in this thesis was kindly provided by a group in KU Leuven, who have utilised this technique for the investigation of facial shape previously. They

suggest itself superiority when investigating correlated landmark data when sample sizes are reduced in comparison to the number of landmarks (Shrimpton et al. 2014; Matthews et al. 2016; Muggli et al. 2017).

With regards to descriptive statistics, score plots associated with PLSR are more difficult to interpret than those possible with conventional PCA and mPCA as the component consider both the variables and the landmark data simultaneously. The PLSR used in this thesis did not utilise a multilevel structure. The benefits of separating the variables at different levels of the model is worth exploring. Initial work has begun, comparing multilevel PLSR with mPCA (Farnell et al. 2021).

It may also be useful to investigate the influence of non-normality in PLSR. Goodhue et al. (2012) suggests that PLSR performs in a similar manner to multiple regression and was relatively robust for non-normally distributed data. Stahle and Wold (1990) also suggest non-normal data is unlikely to be an issue with PLSR. However, this would be worthy of further investigation in future.

7.2.5 Multilevel Principal Component Analysis

mPCA is a useful extension of conventional PCA as it allows data reduction to be conducted in a "guided" manner, considering the group the subjects belong to (e.g., Croatian, English, Welsh or Finnish, male or female). This is advantageous as it is possible to be more confident that the variation seen in the score plots is due to the variable of interest. It is also less likely that variation due to the variable of interest is discarded in the dimension reduction process. As mPCA finds the maximum variation between the groups, it can model subtle differences in the facial shapes. This is advantageous as differences in facial shape are likely to be subtle given that it is not possible to determine (for example) a person's asthma status in everyday life from their facial shape. Only the variables that explained 5% of the total variation in their respective mPCA models reached global significance at a 0.05 level in the PLSR models. It could therefore be suggested that mPCA may be more sensitive than PLSR in determining subtle differences in facial shape and attributing these to a particular variable. However, it may therefore follow that mPCA is prone to overfitting, as by nature, it maximises the variation in the current data set.

The interpretation of the component score plots is easier than in PLSR and conventional PCA. It is therefore useful for data visualisation in the exploratory data analysis phase when data is naturally categorical with equal numbers of subjects in each group. Given that the score plots for mPCA and DFA show very similar patterns of group separation, there is evidence to support the results attributed to mPCA. However, as there are still multiple plots when there is more than one group, determining the true differences between the groups is more challenging than using the raw landmarks in isolation. Furthermore, as was discussed in the results section, where there is a large discrepancy in the group sample sizes, there is a suspicion that the separation in the component score means is artificial. This was also suggested by Bookstein (2019) with regards to between-groups variation, and therefore the score plots should be interpreted with caution at this stage. A further disadvantage of mPCA is the requirement for categorical data. This cannot be overcome directly, but alternative techniques that are capable of handling continuous variables, such as PLSR, can be adapted to utilise a multilevel structure (Farnell et al. 2021). Finally, a large range of component scores was evident in the mPCA models. This was further exaggerated when the scores were standardised. Constraining the component scores generated by the mPCA model to 3SD, as suggested by Farnell (2017), may be of benefit in future.

Given that the eigenvalues take the variables into account in mPCA, it is possible to calculate the percentage of the total variation each variable explains. This is more useful than the information available in conventional PCA. For example, knowing that PC1 explains 18.47% of the total variation, is not particularly useful. However, knowing that population (14.55%), sex (9.96%) and all other variation (66.63%) influence facial shape in different amounts, as is possible with mPCA, is very useful information. This allows variables to be compared with regards to their importance in explaining differences between faces and is one of the strengths of mPCA as a descriptive tool. It also demonstrates the variation that has not been explained by the variables in the model and is represented by the within-group variation level. However, as the number of eigenvalues that can be retained in constrained by the number of groups, variation will not be modelled in all possible directions, and information may therefore be lost.

The use of the percentage of total variation as a variable selection tool prior to further analyses, such as PLSR, is enticing. Variable selection is particularly useful in the field of facial shape as there are likely to be many variables that have a subtle influence either individually or in combination. The difficulty therefore lies in finding and quantifying important variables to add to a statistical model. If an important variable is missed from the analysis, the results could be misleading. If too many unimportant variables are added, there is a risk of overfitting (Tabachnick and Fidell 1996). The output from mPCA is easy to understand and visualise thus making these decisions more intuitive from a clinician's point of view. However, many other methods have been suggested in the literature with regards to variable selection in PLSR. Some of these methods were initially reviewed by Mehmood et al. (2012) and have subsequently been compared by the authors Mehmood et al. (2020). The authors grouped the potential variable selection methods into three categories: filter methods, wrapper methods and embedded methods. Filter methods involve running the PLS model with all possible variables and using a threshold to decide which variables to keep in subsequent models. Wrapper methods use a filter method first, then involve re-running the PLSR model until the best combination of variables and model fit is found. The embedded methods involve modified PLSR algorithms which have the variable selection step "embedded" in the model. Using mPCA as a screening tool could be seen as a type of filter method.

8 LIMITATIONS AND FURTHER WORK

8.1 CLINICAL DATA

The differences explored here with regards to geographical location are across both sexes and with regards to sex are across all geographical locations. This helped to keep the methods comparable and ensured that the results were as clear as possible. Further work on the different levels of sexual dimorphism within each population would be of interest and assessment of whether the differences are similar for both females and males.

BMI was used as a proxy for body fat (visceral obesity). However, BMI has been criticised as it does not distinguish between bone mass, muscle mass and excess fat. It also does not consider age and gender, although some effort has been made to account for ethnicity by adapting the cut offs for obesity (WHO Expert Consultation 2004). There is also debate around whether it correlates well with ill-health (Mahadevan and Ali 2016). However, Lawlor et al. (2010) suggest that in the ALSPAC sample, waist circumference and fat mass measured via x-ray absorbiometry are no more associated with cardiovascular risk factors than BMI. The use of alternative measures would be of use in further work to improve confidence in the results and improve understanding of the impact of body fat on facial shape. Alternative methods include waist circumference (Lee and Kim 2014), dual energy x-ray absorptiometry to assess fat mass (Lawlor et al. 2010), a standardised reference model based on upper arm length and sitting height (Bagust and Walley 2000) and the Relative Fat Mass Index (Woolcott and Bergman 2018):

$$\text{Men} = 64 - \left(\frac{20 * height}{waist\ circumference}\right) \qquad \text{Women} = 76 - \left(\frac{20 * height}{waist\ circumference}\right)$$

The metabolic factors assessed here are snap shots of their levels at one point in time, close to the time of acquisition of the facial images. However, facial shape is an accumulative measure and levels of these metabolic factors are likely to fluctuate over time. It is therefore difficult to draw any conclusive evidence with regards to their influence on facial shape, although it can be used as a guide for further investigation.

Although subjects were excluded if they had obvious craniofacial dysmorphia, further information on any history of trauma, family history of craniofacial issues and the orthodontic status of the subjects would be beneficial. It would also be useful to determine whether the subjects presented with hypodontia or any other dental anomalies. Further information would also be useful with regards to the medications taken by the subjects. In particular, the use of insulin or inhalers. Detailed medical histories would also determine if any other variables should be investigated and allow for more thorough investigation with regards to the severity of the breathing disorders.

Furthermore, several of the variables (pubertal stage, maternal smoking levels and maternal alcohol consumption) were self-declared via questionnaire. It is therefore possible that the levels documented were not accurate. The literature on the reliability of self-reported smoking statuses is conflicting. There is a possibility that there is geographical variation in the reliability of mothers to accurately report smoking levels. In a Norwegian population, Kvalvik et al. (2012) found a sensitivity of 82% and specificity of 92% when comparing self-reported smoking and plasma cotinine levels, concluding that self-reporting is an accurate method of recoding smoking status. Furthermore, a study using a historical population from the 1950s and 60s found that 94.9% of woman who reported they were non-smokers and 87% of smokers reported their smoking status correctly (Klebanoff et al. 1998). A more recent study from Scotland found that self-reported smoking levels were not accurate and accounted for a failure to detect over 2400 smokers (Shipton 2009). A further study on a Hungarian population found that the smoking status reported by the mothers was not consistent with information gained about their smoking status from family members (Czeizel et al. 2004). In future, it may therefore be more accurate to use plasma cotinine levels rather than self-reported smoking levels. The maternal plasma cotinine levels are available for the ALSPAC population for the first trimester (Taylor et al. 2014), thus this could be interesting to explore in future. However, the use of prospective cotinine levels can also include biases as mothers can reduce their smoking levels prior to the blood test (Blood-Siegfried and Rende 2010).

8.2 3D FACIAL IMAGING AND INTERPRETING THE FACIAL DIFFERENCES

8.2.1 Superimposition and Generalised Procrustes Analysis

GPA minimises the sum of the squared distances between the landmarks (Klingenberg 2021). The use of GPA is accepted due to its easy implementation. However, any processing step used on the facial meshes inherently changes the original face and thus can influence the results. One of the disadvantages of GPA is described as the "Pinocchio effect". If one or a few landmarks differ greatly in shape A compared to the other shapes assessed, the centroid of shape A will be influenced. This will subsequently affect the position of the other landmarks after GPA. Therefore, the variation visualised may not represent true facial differences (Klingenberg 2021). Klingenberg (2021) extend the caution required when interpreting facial differences after GPA as the issue of landmarks being influenced by each other is not isolated to situations where a large difference is seen in a small number of landmarks.

In cephalometry, well defined stable structures are used to superimpose images. In surface facial shape research, these structures are not available. It has previously been suggested that the faces could be aligned initially on mid-endocanthion during the superimposition process (Zhurov et al. 2005). This has been proposed as a stable structure as it is in the region of the cribriform plate, which is defined as a stable structure by Björk (1955) due to its position on the anterior cranial base. However, proceeding with GPA makes this step redundant.

The lack of a stable structure or agreed reference point(s) can make it challenging to fully appreciate the effects each of the variables are having as the differences in the facial shapes are all relative to each other. For example, the finding that BMI produces a more retrusive mid-face may in fact be simply that the midface is retrusive in comparison to the more prominent cheeks, rather than being truly retrusive. This may explain some of the disagreement around whether the midface is retrusive or pronounced in previous studies (Ohrn et al. 2002; Ferrario et al. 2004). One method that may reduce this issue would be to segment the facial mesh and analyse each facial structure individually. This is demonstrated by Roosenboom et al. (2018).

8.2.2 Scaled data only

Scaling was carried out during the GPA registration to isolate shape from size. This allows facial shapes to be investigated together despite known facial size differences (e.g., male and female faces could be investigated together). However, this process reduces the information available on the differences between the faces. The biggest disadvantage of scaling the faces is the difficulty in determining the magnitude of the differences between the faces in this thesis. Instead, the differences are relative, and provide information on the facial features that are more likely to change in shape to a greater or lesser extent with various variables. The information in this thesis should therefore be used to inform further analyses when size is not removed and will aid with disentangling whether a variable influences size, shape or both.

8.2.3 Interpreting the differences in the landmarks

When interpreting the differences in the 21 landmarks, it was easy to disentangle the influences in *x*-, *y*- and *z*- directions so that these could be explored separately. This was not so straightforward with the regards to the full facial mesh. Therefore, the differences explored in the latter two sections explore the facial features that are most likely to be influenced by each of the variables. Further work is required to determine the direction of these differences. Furthermore, the distances between landmarks, rather than the positions of the landmarks in isolation, could be explored in further detail. Rather than exploring Euclidean distances (i.e., the straight-line distance between two points), it may be prudent to explore Geodesic distances (the shortest distance along the curved surface of the face) (Abbas et al. 2018).

8.2.4 Soft tissue only

Investigation of facial shape using 3D facial imaging gives information on soft tissue features only. It is possible to estimate differences in hard tissues from these images. However, it is not possible to determine the influence of variables on hard tissues with certainty. As was discussed previously, investigation of hard tissues is invasive and either presents a radiation risk or involves an MRI scan which is expensive, time consuming and cooperation dependent (Greene et al. 2016).

8.3 MULTILEVEL PRINCIPAL COMPONENT ANALYSIS

8.3.1 Discretisation

There are many methods that can be utilised to undertake discretisation (Liu et al. 2002). In this thesis, the subjects with pathological levels of the variable of interest were separated from those presenting with non-pathological levels. Groupings were then chosen within the non-pathological levels, attempting to keep the width of the groups equal. This allowed the discretisation process to be flexible, to gain as much meaningful information from the data whilst maintaining groups with reasonable sample sizes. However, the process is subjective, and the number of groups chosen for each analysis may have influenced the overall importance of each variable. This is particularly salient for mPCA where the number of possible eigenvalues is restrained by the number of groups (the higher the number of groups, the more eigenvalues can be calculated and retained).

The influence of the number of groups was investigated using separate two-level mPCA models on maternal smoking during the first trimester. As Figure 68 shows, as the number of groups increases, the amount of variation attributed to maternal smoking also increases. Where the results of mPCA are used for screening and variable selection, the number of groups chosen could therefore result in potentially important variables being discarded prior to further analysis or increasing the relative importance of unimportant variables. Investigating different discretisation methods and their impact on the results would be of benefit in the future.

With regards to continuous data, avoiding the requirement for discretisation completely would valuable. PLSR can deal with both categorical and continuous data. Preliminary work has begun on the development of multilevel PLSR, which avoids the need to discretise continuous variables, whilst maintaining the benefits of a multilevel model (Farnell et al. 2019; Farnell et al. 2021). An alternative method could be to use groupings whereby only one subject or very few subjects belong to each group. Further work on this, including the impact on the component scores where group sample sizes are very small, would be beneficial.

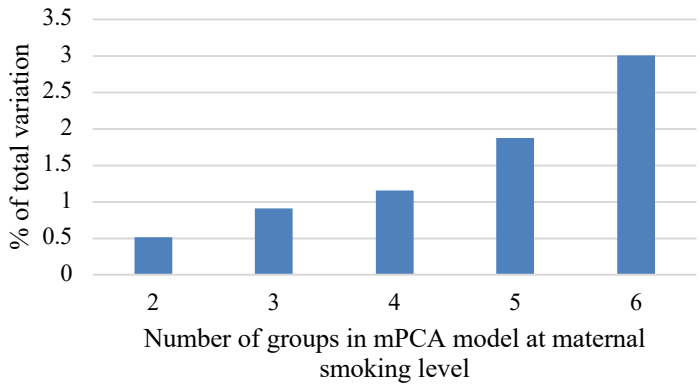


Figure 68: The influence of increasing the number of groups on the total variation explained by the variable. The influence of maternal smoking during the 1st trimester on 21 facial landmarks was used as an example.

8.3.2 Imbalanced data sets

Following discretisation, a clear imbalance of the data set was evident for many of the variables. Meaning, the sample sizes in each category (or group) were different. This is common in data sets. For normally distributed ordinal or continuous data, by definition, there will always be a category with less subjects at both extremes if the categories are of equal width. If the data is skewed, there will always be fewer subjects in the group(s) at the tail of the distribution (Figure 69). For categorical data, in a longitudinal population study such as ALSPAC, it is far less common for a subject to present with a condition than without a condition (e.g., atopy) or present with pathological levels of a variable than non-pathological levels (e.g., fasting insulin levels). Imbalanced data sets are therefore common in studies.

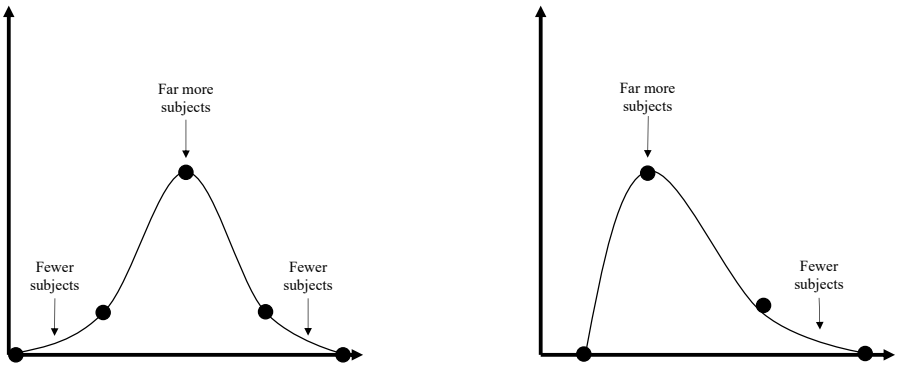


Figure 69: Visualisation of why there will always be fewer subjects at the extremes in both normally distributed and skewed variable data.

Imbalanced data sets are an issue for many analytical techniques. Bookstein (2019) showed that the smallest group is often falsely placed at the end of the axis when visualising the component score plots using between-group PCA. This also appears to happen in mPCA as visualised in study two. It may be that the arguments of Bookstein (2019) hold true for mPCA. This introduces uncertainty around whether the component score plots show real differences or whether the apparent separation of the group means is an anomaly due to differential sample sizes.

Imbalanced data sets may also influence the amount of variation attributed to a particular variable. This was investigated by running separate two-level mPCA models for the 21-landmark data in two ways. Firstly, the random groups were evenly weighted. Subsequently, the random groups were generated so that they were weighted to mimic the sample sizes seen in the maternal smoking during the 1st trimester. In both analyses, the amount of variation explained increased with the number of groups, but this was much more obvious when the group sample sizes differed (Figure 70).

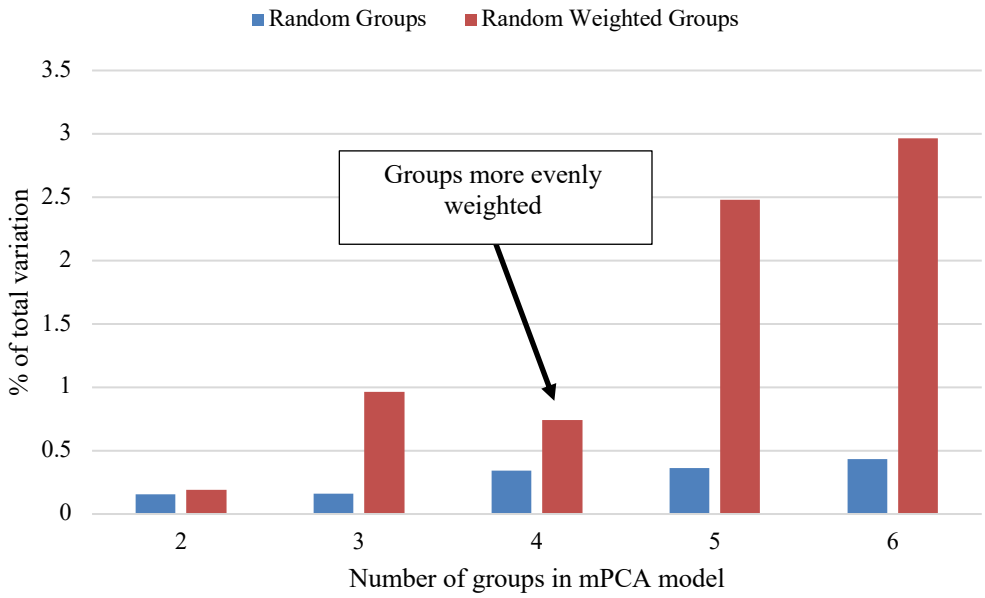


Figure 70: The influence of increasing the number of groups in mPCA in randomly grouped landmark data for both equally weighted group sample sizes and imbalanced groupings on the amount of variation explained.

Within the field of machine learning, it is important that data sets are balanced so that models are not primarily trained on one group. If this was to happen, the model would be excellent at classifying subjects in the majority class (the group with many more subjects) but will likely struggle with the minority class (group with far fewer subjects) (Jain et al. 2017). Two broad methods for balancing the sample sizes of the groups have been suggested for machine learning problems and include over-sampling and under-sampling (Haque et al. 2014). Over-sampling involves using mathematical techniques to artificially generate data so that the data available for the smallest group increases to the size of the larger group. Under-sampling involves randomly removing data from the group with the most data, so the groups become equal in size (Figure 71). One of the distinct disadvantages of the concept of under-sampling is the loss of data.

Methods of over-sampling include random oversampling, Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN). Random oversampling is a simple method which involves randomly replicating data from the minority class. Although simple to execute, this method risks overfitting. SMOTE was first proposed by Chawla et al. (2002) to reduce the chance of overfitting. ADASYN was suggested by Haibo et al. (2008) as an improvement on SMOTE. SMOTE projects lines between the real data points and randomly generates data points along these lines. ADASYN uses a weighted distribution to ensure that the generated data points represent the original data more accurately (Haibo et al. 2008). Preliminary work has begun on the use of random undersampling and ADASYN prior to mPCA.

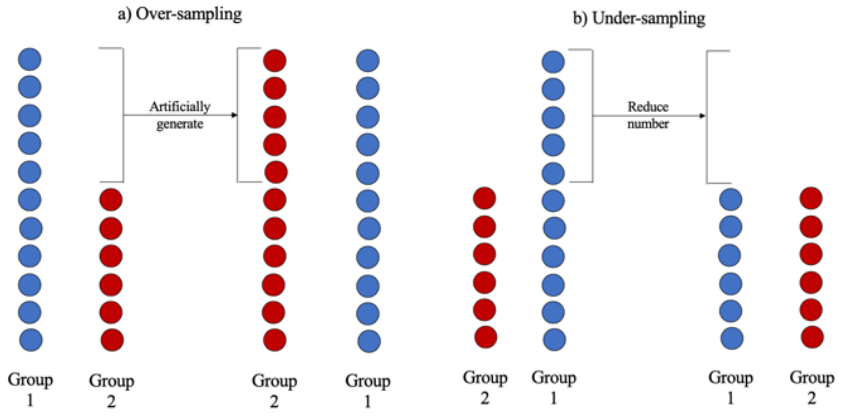


Figure 71: Visualisation of (a) oversampling and (b) undersampling.

8.3.3 Overfitting

Overfitting is a concept that describes a situation where a model fits the data it was generated with very well, but it does not generalise to other data due to the inclusion of a large number of variables (Tabachnick and Fidell 1996). For example, if a further English sample was assessed using the mPCA models developed here using the ALSPAC cohort, the model may not represent the new data well. However, as the mPCA methodology is linear in nature (uses straight lines), it is likely to be more generalisable than a non-linear model. A method of overcoming overfitting is to split the data set into "training" and "test" sets. Subjects in the training set are used to build the model, whilst subjects in the test set are afterwards to determine how good the model is with new data (Xu and Goodacre 2018). The model can be given a success rate, in the form of percentage of the number of the new subjects are classified correctly. Further work on the use of these would be beneficial in ensuring that overfitting is minimised so that mPCA can be used with more confidence.

8.3.4 Number of eigenvalues retained

Selecting the number of eigenvalues to retain in the analyses in this thesis involved visualising the eigenvalue magnitudes on an eigenvalue plot and discarding those deemed small. This is based on Cattell's scree test for factor analysis (Cattell 1966). This is an efficient method for choosing which eigenvalues to keep and ensures that most of the variation between the faces is retained in the analysis. However, it is subjective. Alternative methods include the Kaiser-Guttman method (Guttman 1954), which retains eigenvectors with an eigenvalue above the eigenvalue that explains the mean amount of variation. This was utilised by Djordjevic et al. (2013b) when using conventional PCA and multivariate regression to assess the influence of metabolic factors on facial shape, and by Toma et al. (2012) when assessing facial variation in the ALSPAC population. However, this method excludes many eigenvalues and thus may reduce the opportunity of visualising more subtle differences in facial shape. A further possible method involves retaining the PCs that have an eigenvalue that explains 95% or 99% of the total variation in the data. Although this method ensures that the variation captured in the analyses is standardised, as it does not consider differences in the magnitude of the individual eigenvalues, there is the possibility of including eigenvectors that are noise within the data or excluding eigenvectors that

include meaningful data. With regards to PLSR, cross-validation is a method by which a decision can be made on which eigenvalues to retain. Work on the use of this in mPCA could also be beneficial.

The number of eigenvalues that can be retained in the levels assessing a particular variable (i.e., the between-groups variation levels) is restricted to the number of groups minus one. In this thesis, this did not appear to be an issue for the levels assessing between-groups variation where groups were equal to or exceeded three, as it was rarely required to retain more than two PCs for the variables. Where there are only two groups, only one eigenvector can be retained (e.g., sex). This has the potential to reduce the amount of information available for analysis as the data set is only analysed in one direction (PC1). There is no clear solution for this issue and is likely to be one of the true disadvantages of the technique. However, with regards to continuous data, using very small sample sizes in each grouping so that the mean or median of each group would represent either one individual or a small number of individuals with very similar levels of a continuous variable (as discussed previously) may not only reduce issues associated with discretisation, but also eliminate the restrictive nature of the number of eigenvalues that can be retained. Further work in this area would be beneficial.

Furthermore, it was challenging to determine how many eigenvalues to retain at the within-group variation level. Preliminary investigations surrounding the number of eigenvalues to retain at this level suggested that the component scores of the subjects at the extremes are particularly influenced by the number of retained eigenvectors. However, these investigations also suggest that the number of retained eigenvectors does not influence the mean component scores for each group centroids and it is therefore unlikely that this would influence the results of this thesis. Further work in this area would be beneficial.

8.3.5 Inference

Inference was determined for the mPCA models via ANOVA or MANOVA. Subsequent bootstrapping of the component scores was carried out, based on a baseline F-statistic determined via ANOVA or MANOVA. Bootstrapping provides a

more robust estimate of *p*-values when data is non-normally distributed and / or outliers occur. Further work on post-hoc tests to determine which groups were significantly different to each other could be of benefit, particularly where it is not clear which group is most differentiable from the others.

8.3.6 Number of landmarks and computational time

Unfortunately, in its current form, it is not possible to use the full 7160 three-dimensional landmarks in the mPCA code as the required matrix exceeds the maximum matrix size in MATLAB. Adapting the code and/or increasing computing power to circumvent this limitation would be beneficial. With regards to computational time, generating the eigenvalues and eigenvectors, and thus determining the relative importance of each variable takes seconds. This makes mPCA an efficient screening tool for variable selection prior to further analysis. It is also quick to calculate the influence of these eigenvectors and eigenvalues on the individual landmarks. However, calculating the component scores iteratively using the Global Optimiser can take hours depending on the number of subjects and landmarks as well as the number of eigenvalues retained. To make mPCA an accessible and useful data exploration tool for as many researchers as possible, without access to high powered computer hardware, improvements in this area would be of benefit. The first step is likely to utilise MATLAB's parallel toolkit.

8.3.7 Limitations of number of levels, missing data and lack of variable adjustment

Although mPCA separates the between and within group variation, variables are not explicitly adjusted for unless they are added as different levels in the model. This is suggested in study two where sex, height, BMI and fasting insulin all showed the same result, but on slightly different scales. Unfortunately, it is not possible to build such a model with all four variables as the sample sizes in the resulting categories become too small and the mPCA model is unable to handle groupings with zero subjects. Adapting the mPCA methodology to allow for groups with zero subjects and being able to handle missing data would significantly help extend the number of levels

possible. Also, perhaps a preliminary step could also be used to adjust for variables prior to using the mPCA model.

8.3.8 Further alternative techniques for comparison

It was felt that comparing mPCA to using the raw landmarks only, conventional PCA and DFA provided enough information to ascertain the relative advantages and disadvantages of each technique. It was also hoped that presenting mPCA alongside PLSR demonstrated a use for mPCA within the data analysis toolkit. However, there are many more techniques that could be compared to mPCA; for example, factor analysis, t-distributed stochastic neighbor embedding (t-SNE), support vector machines (SVM), Bootstrapped Response Imputation Modelling (BRIM), decision trees, random forests and auto-encoders (Yong and Pearce 2013; Claes et al. 2014; Wang et al. 2016; Zhou et al. 2018; Chao et al. 2019; Reddy et al. 2020). Furthermore, there was a focus on frequentist statistics in this thesis. The use of Bayesian statistics, which consider uncertainty, would also be useful to explore in future (Ashby 2006).

Of particular interest, could be random forest, given that the technique is easy for clinicians to understand and interpret (Kotu and Deshpande 2015). The ability to understand the steps of a chosen technique is essential from a clinical perspective as the aim is to understand differences between facial shapes (as is the case in this thesis). This contrasts with prediction problems of a non-clinical nature, where understanding the architecture of a model may not be as important. Random forest is a technique that uses a collection of decision trees (Oshiro et al. 2012). Decision trees are visually comparable to an upside tree (Kotu and Deshpande 2015). They start with a base and propose classification options like the branches of a tree (Figure 72). In random forest, each decision tree decides which group a subject should belong to, these are compared and the group with the most “votes” is chosen as the answer (Figure 73).

One of the criticisms of decision trees and random forest is that they can be prone to overfitting (fits the current data well but does not perform well with new data) (Kotu and Deshpande 2015). This is a common theme with regards to modelling facial shape and has been raised as a potential issue with all the techniques discussed in this thesis. It is therefore important to use appropriate techniques to mitigate against this. A

method of reducing this is known as pruning, whereby sections of the model that are deemed unimportant are removed (Kotu and Deshpande 2015). Exploring the use of mPCA in aiding this process would be worthwhile in the future.

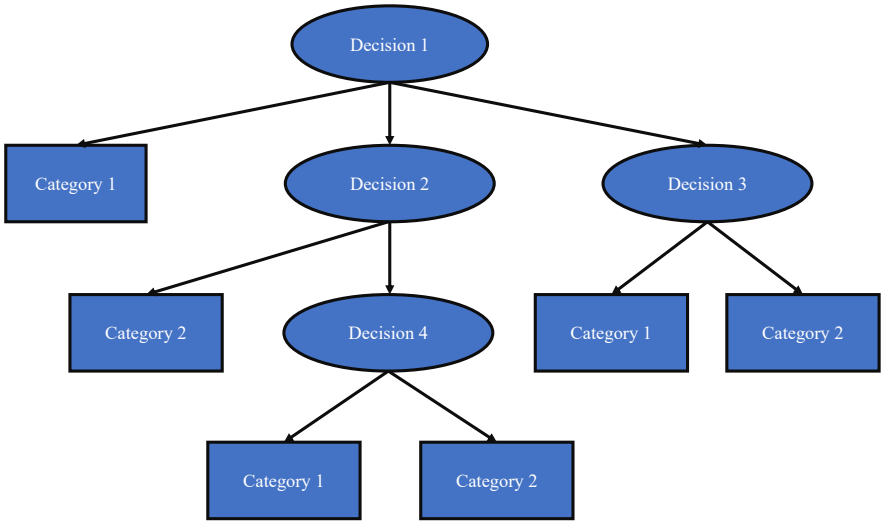


Figure 72: Diagram of a decision tree

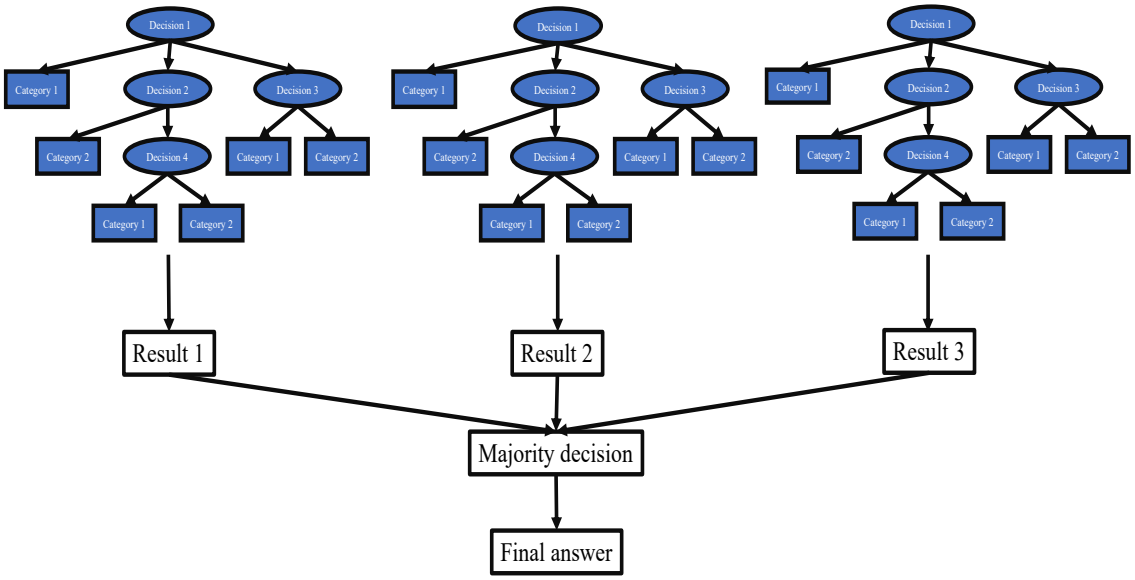


Figure 73: Diagram of a random forest

However, compared to decision trees and random forest, deep learning neural networks are likely to be more powerful. Indeed, a study by AbdAlmageed et al. (2020) compared a deep learning model to DFA and random forest in their ability to determine facial differences between control subjects and those with Congenital Adrenal Hyperplasia (CAH). The deep learning model significantly out-performed DFA and random forest. They found that the nose and upper face were most distinguishable in CAH patients. Their deep learning model predicted the correct grouping 92% of the time, outperforming DFA (86%) and random forest (83%).

A clear disadvantage of deep learning models at the time of writing is the difficulty in interpreting the underlying structure of the model (Montavon et al. 2018). The decisions made by deep neural networks are not explicitly obvious (described as a black box), so it is challenging to know how the algorithm has come to its decision. It is therefore difficult to understand why faces are different from each other, only that the algorithm can differentiate between them. Further information on understanding and interpreting neural networks is discussed by Montavon et al. (2018).

As this field progresses further, it is likely that our understanding of facial shape differences will improve rapidly. Given the complex nature of facial shape and the expanding number of variables associated with facial features, mPCA in isolation is unlikely to be able to disentangle all the necessary variables associated with facial shape due to the practical limitations on the number of levels possible. More sophisticated techniques in the machine learning field may be more successful. However, caution is required to ensure that robust data collection and acknowledgement of variables is maintained rather than relying solely on progressively more complex algorithms. The use of mPCA as a preliminary step to assess the relative importance of the variables introduced to deep learning models are known prior to building the model may help in their clinical interpretation.

8.4 CLINICAL RELEVANCE AND RELATED FURTHER WORK

Translating the work in this thesis into improvements in patient care can be discussed with regards to short-term possibilities and more far-reaching, long-term possibilities. In the immediate future, the results of thesis can be used to determine which confounding factors to consider in further research within orthodontics, forensic anthropology, or psychology. The mPCA results, confirmed by the PLSR results, suggest that sex, height, and BMI should be taken into account in all facial shape studies. The results with regards to fasting insulin also suggest that this should also be considered as a potential confounding factor, particularly where the mandible and brow-ridge are concerned.

In the near future, the development and implementation of prediction models could help to aid orthodontic treatment planning decisions. As discussed at the beginning of this thesis, one area that could benefit is helping to predict the potential success rates of treatments, such as twin block appliances. Although it is important that patients are not declined treatment due to an algorithm in isolation, prediction models could be used in conjunction with clinical expertise to optimise the timing of treatment commencement or help the clinician in determining when a successful outcome is no longer likely. A further use of prediction models, which has been investigated extensively, is growth. Again, as discussed previously, of particular use in the orthodontics is predicting the growth potential in class III patients to help determine whether a case will be camouflageable or require surgery.

To build reliable prediction models, it is important that models contain as much information as possible whilst limiting the risk of overfitting by excluding redundant variables (Tabachnick and Fidell 1996). mPCA could provide a method for screening variables prior to building complex models to ensure that the most important variables are included. The work in this thesis has started this process. Again, the results of this thesis suggest that sex, height, and BMI would be essential inclusions in any prediction models. This is to be expected. However, of particular interest is the relatively high variation explained by fasting insulin. This is further emphasised by the possibility of increased levels of insulin restricting mandibular growth, particularly in males.

It could therefore be suggested that screening patient insulin levels may provide useful information on the clinical variability seen with regards to treatment success and growth. However, further work is required to determine the long-term influence of insulin and to determine how insulin levels fluctuate in vivo, throughout childhood and into adolescence. This information, in conjunction with skeletal pattern and severity, would provide more robust information than the current cross-sectional study. Moreover, given that insulin appears to influence GH levels differently in obese individuals compared to those without within a healthy BMI range (Qui et al. 2017), it would also be useful to stratify individuals by BMI to investigate this with relation to facial shape. Further work into the potential for shared genetics will also be worthwhile. Continuing to investigate females and males separately would be useful given the differing influence of insulin in this thesis. Non-linear models could also be explored. However, to assess these areas reliably, sample sizes need to be increased as well as utilising sample size calculations to mitigate type II errors. The results of this thesis could help in any future sample size calculations. Combining datasets and increasing dataset sizes by pooling resources nationally and internationally, will help to make further work more reliable.

More far reaching is the possibility of interfering with the natural growth of patients via manipulation of growth factor levels, dietary changes or genetic modification, under the umbrella of personalised orthodontics. Indeed, others have suggested that growth factors could be used to manipulate craniofacial shape and mitigate the need for active appliance therapy (Barton and Crowder 2010). Again, of particular interest is the possible influence of insulin levels on the mandible. Perhaps the reduction of insulin levels through dietary changes could reduce the incidence or severity of class II malocclusions. From a service perspective, this in turn could reduce the requirement for treatment and thus eliminate or reduce the direct and indirect costs associated with orthodontic intervention. Furthermore, given the documented benefits in ensuring optimal insulin levels for general health (Rahman et al. 2021), the possibility of improving facial harmony may serve to enhance public health messages.

However, it is important that the variables that explained less than 5% of the variation in their respective models are not completely discounted. Firstly, this thesis focuses on the influence on the total facial shape, whilst orthodontics primarily centres around

the maxilla and mandible. Segmentation of the facial shells and repeating the analyses here would help to determine the factors most relevant to orthodontics. Secondly, when the variation explained by these variables is taken accumulatively, the total variation may lead to clinically significant differences in facial shape. In a similar manner to the building of the PLSR models in this thesis, alternative prediction models could be investigated by adding and removing the variables assessed here prior to determining the variables that are of particular importance. Again, increasing sample sizes will help mitigate type II errors and help to determine whether variables are of significance. This research would also build on the work in this thesis to help determine the usefulness of mPCA as a variable selection method. It would also be interesting to explore the magnitude of facial shape differences that lead to clinically distinguishable differences.

Out-with orthodontics, there may also be benefits to exploring the influence of insulin further. Facial shape could be used as a screening measure for increased insulin levels in childhood (particularly if these increased levels are chronic in nature). This is less invasive than a blood test and could potentially be used in combination with measures of body fat. Furthermore, knowledge of the effect of insulin on the brow-ridge could be used to improve identification of sex. A prominent brow-ridge is traditionally associated with males (for example, (Ferrario et al. 2003; Kau et al. 2006; Bozic et al. 2009; Gor et al. 2010)). However, differing insulin levels may explain some of the overlap between the sexes with regards to this facial feature given that female subjects appear to have more prominent brow-ridges in the presence of increased insulin.

9 CONCLUSIONS

9.1 CLINICAL FINDINGS

Strong evidence for association with adolescent facial shape

Null hypothesis rejected with all tests following Bonferroni correction
Explain greater than 5% of variation in mPCA model(s)

- **Geographical location**
  - *Tests: PCA & MANOVA, DFA & Wilks Lambda, mPCA (3-levels) & MANOVA*
  - Variation explained in mPCA model (21 landmarks): **11.34 – 14.59%**
  - Facial differences: All facial features influenced
  - Novel findings: Croatians most distinguishable (caution with this result), more detail provided on facial differences than before, agreement between landmark only data and mPCA suggests mPCA useful tool
- **Sex**
  - *Tests: PCA & ANOVA, DFA & Wilks Lambda, mPCA (2-levels) & ANOVA, mPCA (3-levels) & ANOVA*
  - Variation explained in mPCA model: **9.98 – 10.93%** (21 landmarks); **17.31%** (1000 quasi-landmarks)
  - More prominent/ larger (female): Cheeks, infra-orbital regions
  - Less prominent/ smaller (male): Brow ridge, nose, lips, under chin (i.e., rounder face)
  - Novel findings: Variation explained, agreement with previous research and between analyses suggests mPCA useful tool
- **BMI**
  - *Tests: mPCA (2-levels) with ANOVA, PLSR with bootstrapping*
  - Variation explained in mPCA model (1000 quasi-landmarks): **17.29%**
  - More prominent/ larger: Cheeks, supra-orbital region, forehead
  - Less prominent/ smaller: Nose, lips, under chin (i.e., rounder face)
  - Novel findings: Variation explained, agreement with most previous research and between analyses suggests mPCA useful tool
- **Height**
  - *Tests: mPCA (2-levels) with ANOVA, PLSR with bootstrapping*
  - Variation explained in mPCA model (1000 quasi-landmarks): **11.05%**
  - More prominent/ larger: Nose, chin, brow ridge
  - Less prominent/ smaller: Cheeks
  - Novel findings: Variation explained, adolescent facial differences explained for first time

Strong evidence for association with adolescent facial shape with mPCA model and with PLSR models before Bonferroni correction but not after Bonferroni correction

Tests: mPCA (2-levels) with MANOVA (with Bonferroni correction) and PLSR with bootstrapping (with and without Bonferroni correction)
Null hypothesis rejected with mPCA model and PLSR model before Bonferroni correction
Null hypothesis accepted with PLSR model after Bonferroni correction
Explains greater than 5% of variation in mPCA model

- **Fasting insulin**
    - Variation explained in mPCA model (1000 quasi-landmarks): **5.83%**
    - More prominent/ larger: Cheeks (buccal fat pad), brow ridge/ supra-orbital region
    - Less prominent/ smaller: Chin, nose, intra-orbital region
    - Novel findings: Variation explained, more facial differences explained than previous research, first time finding a significant result (mPCA only)

Evidence for association with adolescent facial shape with mPCA models but not with PLSR models

Tests: mPCA (2-levels) with (M)ANOVA and PLSR with bootstrapping (both with Bonferroni correction)
Null hypothesis rejected (mPCA) and accepted (PLSR)
Explain greater than 2% of variation in mPCA model

- **Other metabolic factors (cholesterol, VLDL, triglycerides, LDL)**
    - Variation explained in mPCA models (1000 quasi-landmarks): **3.55%** (cholesterol), 3.54% (VLDL), 2.25% (triglycerides), 2.01% (LDL)
    - Facial differences differ between mPCA and PLSR
    - More prominent/ larger (mPCA): Cheeks, supra-orbital regions
    - Less prominent/ smaller (mPCA): Nose, lips, under chin (i.e., rounder face)
    - Novel findings: Variation explained, significant findings for first time (mPCA only), facial differences explained in detail for first time
- **Maternal smoking (1st trimester, 2nd trimester)**
    - Variation explained in mPCA models (1000 quasi-landmarks): **2.51%** (1st trimester), **3.74%** (2nd trimester)
    - Facial differences differ between mPCA and PLSR
    - More prominent/ larger (mPCA): Outer surface of cheeks
    - Less prominent/ smaller (mPCA): Nose

- Novel findings: Variation explained, significant findings for first time in non-syndromic/ non-cleft adolescents (mPCA only), adolescent facial differences explained in detail for first time

- **Maternal alcohol consumption (1st trimester)**
  - Variation explained in mPCA models (1000 quasi-landmarks): **2.62%** (1st trimester)
  - Facial differences differ between mPCA and PLSR
  - More prominent/ larger (mPCA): Forehead, supra-orbital region, lips
  - Less prominent/ smaller (mPCA): Nose, chin
  - Novel findings: Variation explained, significant findings for first time in non-syndromic adolescents (mPCA only), adolescent facial differences explained in detail for first time

Evidence for association with adolescent facial shape with mPCA models but explain less than 2% of total variation therefore not included in PLSR models

Tests: mPCA (2-levels) with (M)ANOVA
Null hypothesis rejected after mPCA with (M)ANOVA and Bonferroni correction
Explains less than 2% of variation in mPCA model

- **Other metabolic factors (HDL, fasting glucose)**
  - Variation explained in mPCA models (1000 quasi-landmarks): **1.58% (HDL), 1.69% (glucose)**
  - More prominent/ larger (mPCA): Cheeks (HDL), infra-orbital region (HDL and low glucose), nasal tip (HDL), forehead (low glucose)
  - Less prominent/ smaller (mPCA): Brow ridge (HDL), around mouth (low glucose)
  - Novel findings: Variation explained, significant findings for first time (mPCA only), facial differences explained in detail for first time
- **Pubertal stage (pubic hair, genital development)**
  - Variation explained in mPCA models (1000 quasi-landmarks): **1.76%** (pubic hair), **0.58%** (genital development)
  - More prominent/ larger (mPCA – stage 3 compared to 4/5): Nose, maxillary region, orbits
  - Less prominent/ smaller (mPCA – stage 3 compared to 4/5): Chin (genitals), brow ridge (both)
  - Novel findings: Variation explained, facial differences explained in detail for first time
- **Maternal smoking (before pregnancy)**
  - Variation explained in mPCA models (1000 quasi-landmarks): **1.53%**
  - Unwise to interpret facial differences as component scores not ordered
  - Novel findings: Variation explained, significant findings for first time (mPCA only)
- **Maternal alcohol consumption (before pregnancy, 2nd trimester)**
  - Variation explained in mPCA models (1000 quasi-landmarks): **1.56%** (before), **1.60%** (2nd trimester)

- More prominent/ larger (mPCA): Forehead (both), chin asymmetry (2nd tri), side of nose (2nd tri)
- Less prominent/ smaller (mPCA): Nose tip and bridge (both), infra-orbital regions (before)
- Novel findings: Variation explained, findings for 2nd trimester generally agree with previous research with some new facial features implicated here

- **Breathing disorders (asthma (0-3.5 years old), asthma (7.5 years old), hay fever)**
  - Variation explained in mPCA models (1000 quasi-landmarks):
    - **0.90%** (asthma 0-3.5 years old), **0.40%** (asthma 7.5 years old), **0.35%** (hay fever)
  - More prominent/ larger (mPCA): Nose (hay fever), longer face shape (asthma 0-3.5yo), mandible (asthma 7.5yo), infra-orbital regions (asthma 7.5yo)
  - Less prominent/ smaller (mPCA): Nose (asthma), maxillary sinus/ upper lip region (all)
  - Novel findings: Variation explained, general agreement with previous studies, more facial differences explained here
- **Atopy**
  - Variation explained in mPCA models (1000 quasi-landmarks): **0.15%**
  - More prominent/ larger (mPCA): Nose, forehead, supra-orbital regions, asymmetry of mandible
  - Less prominent/ smaller (mPCA): Maxillary sinus region, lips
  - Novel findings: Variation explained, more facial differences presented here than in previous study

Unlikely to be associated with adolescent facial shape

Tests: mPCA (2-levels) with MANOVA
Null hypothesis accepted after mPCA with MANOVA and Bonferroni correction
Explains less than 2% of variation in mPCA model

- **Age (14 -16 years old)**
  - Variation explained in mPCA model (1000 quasi-landmarks): **1.18%**
  - Novel findings: Variation explained, non-significant result at crucial timepoint during development

9.2 MULTILEVEL PRINCIPAL COMPONENT ANALYSIS

mPCA is a technique that is easy to understand (with some basic background knowledge) both in terms of its methodology and in the graphics it can generate. With regards to categorical variables, mPCA is more useful than conventional PCA in facial shape research. mPCA also appears to be as effective as DFA in separating groups, but with more relaxed sample size constraints. This negates the need for conventional PCA to be conducted first when sample sizes are small in comparison to the number of landmarks and makes interpreting the differences between faces easier with mPCA. When variables are continuous in nature, mPCA is currently most useful as a variable selection tool prior to further analysis. It is therefore valuable as an additional technique in the data exploration phase of data analysis. mPCA may find significant differences between groups more readily than PLSR with bootstrapping. However, caution is required when interpreting the component scores when the data set is imbalanced.

Given the highly complex nature of facial shape, it is unlikely that, in isolation, mPCA will be able to disentangle the effects of all the possible variables reliably. More complex techniques, under the machine learning umbrella are likely to prove more powerful. However, their principal disadvantage at the time of writing is the difficulty in interpreting what the machine learning models are doing. This makes them far less useful from a clinical perspective. In future, the facial shape field requires much larger data sets and the ability to interpret highly complex models to disentangle the influence of genetic and environmental factors on facial shapes reliably.

10 REFERENCES

Abbas, H. et al. 2018. A 3D morphometric perspective for facial gender analysis and classification using geodesic path curvature features. *Computational Visual Media* 4(1), pp. 17-32.

AbdAlmageed, W. et al. 2020. Assessment of Facial Morphologic Features in Patients With Congenital Adrenal Hyperplasia Using Deep Learning. *JAMA Network Open* 3(11), p. e2022199.

Abdi, H. 2011. Partial least square regression. In: Salkind, N.J. ed. *Encyclopedia of measurement and statistics*. Thousand Oaks: SAGE Publications, pp. 741-744.

Abu Alhaija, E. S. and Richardson, A. 2003. Growth prediction in Class III patients using cluster and discriminant function analysis. *European Journal of Orthodontics* 25(6), pp. 599-608.

Adams, M.J. 2005. Chemometrics and statistics: Multivariate classification techniques. In: Worsfold, P. et al. eds. *Encyclopedia of Analytical Science*. 2nd ed. Edinburgh: Elsevier, pp. 21-27.

Afrand, M. et al. 2014. Anterior cranial-base time-related changes: A systematic review. *American Journal of Orthodontics and Dentofacial Orthopedics* 146(1), p. 21-32.e26.

Al Ali, A. et al. 2014a. A three-dimensional analysis of the effect of atopy on face shape. *European Journal of Orthodontics* 36(5), pp. 506-511.

Al Ali, A. et al. 2014b. The influence of asthma on face shape: a three-dimensional study. *European Journal of Orthodontics* 36(4), pp. 373-380.

Al Ali, A. et al. 2015. The influence of snoring, mouth breathing and apnoea on facial morphology in late childhood: a three-dimensional study. *BMJ Open* 5(9), p. e009027.

Alderman, B. W. et al. 1994. Increased risk of craniosynostosis with maternal cigarette smoking during pregnancy. *Teratology* 50(1), pp. 13-18.

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

Allison, D. B. and Saunders, S. E. 2000. Obestity in North America: An overview. *Medical Clinics of North America* 84(2), pp. 305-332.

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

Andersen, M. R. et al. 2004. Reduced endothelial nitric oxide synthase activity and concentration in fetal umbilical veins from maternal cigarette smokers. *American Journal of Obstetrics and Gynecology* 191(1), pp. 346-351.

Andersen, M. R. et al. 2009. Smoking cessation early in pregnancy and birth weight, length, head circumference, and endothelial nitric oxide synthase activity in umbilical and chorionic vessels: an observational study of healthy singleton pregnancies. *Circulation* 119(6), pp. 857-864.

Argente, J. et al. 1997. Multiple Endocrine Abnormalities of the Growth Hormone and Insulin-Like Growth Factor Axis in Prepubertal Children with Exogenous Obesity: Effect of Short- and Long-Term Weight Reduction1. *The Journal of Clinical Endocrinology & Metabolism* 82(7), pp. 2076-2083.

Artese, F. 2019. A broader look at Interceptive Orthodontics: What can we offer? *Dental Press Journal of Orthodontics* 24(5), pp. 7-8.

Ashby, D. 2006. Bayesian statistics in medicine: a 25 year review. *Statistics in Medicine* 25(21), pp. 3589-3631.

Aung, S. C. et al. 1995. Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements. *British Journal of Plastic Surgery* 48(8), pp. 551-558.

Bagust, A. and Walley, T. 2000. An alternative to body mass index for standardizing body weight for stature. *QJM: An International Journal of Medicine* 93(9), pp. 589-596.

Barton, E. R. and Crowder, C. 2010. Growth factor targets for orthodontic treatments. *Seminars in Orthodontics* 16(2), pp. 128-134.

Baumrind, S. and Frantz, R. C. 1971. The reliability of head film measurements. 1. Landmark identification. *American Journal of Orthodontics* 60(2), pp. 111-127.

Baxter, M. J. 1995. Standardization and Transformation in Principal Component Analysis, with Applications to Archaeometry. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 44(4), pp. 513-527.

Baydaş, B. et al. 2007. Heritability of facial proportions and soft-tissue profile characteristics in Turkish Anatolian siblings. *American Journal of Orthodontics and Dentofacial Orthopedics* 131(4), pp. 504-509.

Beaty, T. H. et al. 2013. Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. *Human Genetics* 132(7), pp. 771-781.

Bergeria, C. L. et al. 2017. Comparing Smoking Topography and Subjective Measures of Usual Brand Cigarettes Between Pregnant and Non-Pregnant Smokers. *Nicotine & Tobacco Research* 20(10), pp. 1243-1249.

Birgfeld, C. B. and Heike, C. 2012. Craniofacial microsomia. *Seminars in Plastic Surgery* 26(2), pp. 91-104.

Björk, A. 1955. Cranial base development: A follow-up x-ray study of the individual variation in growth occurring between the ages of 12 and 20 years and its relation to brain case and face development. *American Journal of Orthodontics* 41(3), pp. 198-225.

Bjork, A. 1963. Variations in the growth pattern of the human mandible: longitudinal radiographic study by the implant method. *Journal of Dental Research* 42(1)Pt 2, pp. 400-411.

Björk, A. and Skieller, V. 1983. Normal and abnormal growth of the mandible. A synthesis of longitudinal cephalometric implant studies over a period of 25 years. *European Journal of Orthodontics* 5(1), pp. 1-46.

Blood-Siegfried, J. and Rende, E. K. 2010. The long-term effects of prenatal nicotine exposure on neurologic development. *Journal of Midwifery & Womens Health* 55(2), pp. 143-152.

Bookstein, F. L. 2019. Pathologies of Between-Groups Principal Components Analysis in Geometric Morphometrics. *Evolutionary Biology* 46(4), pp. 271-302.

Boyd, A. et al. 2013. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology* 42(1), pp. 111-127.

Bozic, M. et al. 2009. Facial morphology of Slovenian and Welsh white populations using 3-dimensional imaging. *The Angle Orthodontist* 79(4), pp. 640-645.

Brash, J. C. 1924. The growth of the jaws and palate. *The Growth of Jaws, Normal and Abnormal, in Health and Disease*. London: Dental Board of the United Kingdom, pp. 23-66.

Breh, R. and Kamat, N. V. 2015. Cranial base morphology determining sagittal and vertical relation: a cross sectional study. *International Journal of Science and Research* 6(10), pp. 1823-1826.

Brown, A. E. and Walker, M. 2016. Genetics of Insulin Resistance and the Metabolic Syndrome. *Current cardiology reports* 18(8), pp. 75-75.

Bruin, J. E. et al. 2010. Long-term consequences of fetal and neonatal nicotine exposure: a critical review. *Toxicological sciences : an official journal of the Society of Toxicology* 116(2), pp. 364-374.

Bugaighis, I. et al. 2013. Three-dimensional gender differences in facial form of children in the North East of England. *European Journal of Orthodontics* 35(3), pp. 295-304.

Buppajarntham, S. 2019. *Insulin*. Available at: https://emedicine.medscape.com/article/2089224-overview [Accessed: 22nd April 2021].

Camison, L. et al. 2018. Validation of the Vectra H1 portable three-dimensional photogrammetry system for facial imaging. *International Journal of Oral and Maxillofacial Surgery* 47(3), pp. 403-410.

Carels, C. et al. 2001. A quantitative genetic study of cephalometric variables in twins. *Clinical Orthodontics and Research* 4(3), pp. 130-140.

Carmichael, S. L. et al. 2008. Craniosynostosis and maternal smoking. *Birth Defects Research Part A: Clinical and Molecular Teratology* 82(2), pp. 78-85.

Cattell, R. B. 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research* 1(2), pp. 245-276.

Chao, G. et al. 2019. Recent Advances in Supervised Dimension Reduction: A Survey. *Machine Learning and Knowledge Extraction* 1(1), pp. 341-358.

Chawla, N. V. et al. 2002. Smote: synthetic minority over-sampling technique. . *Journal of Artificial Intelligence Research* 16, pp. 321–357.

Chellappa, R. 2005. Face recognition from still images. In: Chellappa, R. et al. eds. *Recognition of Humans and Their Activities Using Video*. United States: Morgan & Claypool, pp. 9-31.

Chen, F. et al. 2011. Age and sex related measurement of craniofacial soft tissue thickness and nasal profile in the Chinese population. *Forensic science international* 212(1-3), p. 272.e271-276.

Christianson, R. E. 1980. The relationship between maternal smoking and the incidence of congenital anomalies. *American Journal of Epidemiology* 112(5), pp. 684-695.

Claes, P. et al. 2012. Improved facial outcome assessment using a 3D anthropometric mask. *International Journal of Oral and Maxillofacial Surgery* 41(3), pp. 324-330.

Claes, P. et al. 2014. Modeling 3D Facial Shape from DNA. *PLoS genetics* 10(3), p. e1004224.

Claes, P. and Shriver, M. D. 2016. New Entries in the Lottery of Facial GWAS Discovery. *PLoS genetics* 12(8), p. e1006250-e1006250.

Cobourne, M. T. and DiBiase, A. T. 2016. The orthodontic patient: Examination and diagnosis. In: Cobourne, M.T. and DiBiase, A.T. eds. *Handbook of Orthodontics*. 2nd ed. Edinburgh: Elsevier, pp. 163-226.

Çokakoğlu, S. et al. 2016. Do Different Orthodontic Malocclusions Affect Patients' Self-Concept and Psychosocial Status? *Turkish Journal of Orthodontics* 29(2), pp. 27-30.

Cole, J. B. et al. 2016. Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. *PLoS genetics* 12(8), p. e1006174.

Coleman, T. et al. 2012. A Randomized Trial of Nicotine-Replacement Therapy Patches in Pregnancy. *New England Journal of Medicine* 366(9), pp. 808-818.

Collett, A. R. and West, V. C. 1993. Terminology of facial morphology in the vertical dimension. *Australian Dental Journal* 38(3), pp. 204-209.

Colvin, J. S. et al. 1996. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nature Genetics* 12(4), pp. 390-397.

Commission, E. 2004. *European guidelines on radiation protection in dental radiography. The safe use of radiographs in dental practice*. Available at: https://ec.europa.eu/energy/sites/ener/files/documents/136.pdf [Accessed: 20th April 2021].

Copray, J. C. 1986. Growth of the nasal septal cartilage of the rat in vitro. *Journal of Anatomy* 144, pp. 99-111.

Copray, J. C. and Duterloo, H. S. 1986. A comparative study on the growth of craniofacial cartilages in vitro. *European Journal of Orthodontics* 8(3), pp. 157-166.

Cousminer, D. L. et al. 2013. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Human Molecular Genetics* 22(13), pp. 2735-2747.

Cousminer, D. L. et al. 2018. First Genome-Wide Association Study of Latent Autoimmune Diabetes in Adults Reveals Novel Insights Linking Immune and Metabolic Diabetes. *Diabetes Care* 41(11), pp. 2396-2403.

Czeizel, A. E. et al. 2004. Smoking and alcohol drinking during pregnancy. The reliability of retrospective maternal self-reported information. *Central European Journal of Public Health* 12(4), pp. 179-183.

De Jong, S. 1993. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* 18(3), pp. 251-263.

De Moor, P. et al. 1972. Growth hormone and the steroid binding β-globulin of human plasma. *Journal of Steroid Biochemistry* 3(3), pp. 593-600.

Demayo, C. et al. 2009. Face Shapes Of Diabetics And Non-Diabetics Described Using Geometric Morphometrics. *Biology* 6(1), pp. 1-6.

Desai, B. B. 2000. *Handbook of Nutrition and Diet*. New York: Marcel Dekker, Inc.

Dibbets, J. M. H. 1996. Morphological associations between the Angle classes. *European Journal of Orthodontics* 18, pp. 111-118.

Djordjevic, J. et al. 2013a. Three-dimensional analysis of facial shape and symmetry in twins using laser surface scanning. *Orthodontics & Craniofacial Research* 16(3), pp. 146-160.

Djordjevic, J. et al. 2013b. A population-based cross-sectional study of the association between facial morphology and cardiometabolic risk factors in adolescence. *BMJ Open* 3(5), p. e002910.

Djordjevic, J. et al. 2016. Genetic and Environmental Contributions to Facial Morphological Variation: A 3D Population-Based Twin Study. *PLoS One* 11(9), p. e0162250.

Editor. 2019. *Fasting blood sugar levels*. Available at: https://www.diabetes.co.uk/diabetes_care/fasting-blood-sugar-levels.html [Accessed: 22nd April 2021].

Egger, G. et al. 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429(6990), pp. 457-463.

Einarson, A. and Riordan, S. 2009. Smoking in pregnancy and lactation: a review of risks and cessation strategies. *European Journal of Clinical Pharmacology* 65(4), pp. 325-330.

El-Mehallawi, I. H. and Soliman, E. M. 2001. Ultrasonic assessment of facial soft tissue thicknesses in adult Egyptians. *Forensic Science International* 117(1-2), pp. 99-107.

Emmanuel, M. and Bokor, B. R. 2020. *Tanner stages*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK470280/ [Accessed: 22nd April 2021].

Enlow, D. H. 1990. *Facial growth*. 3rd ed. Philadelphia: W. B. Saunders.

Farkas, L. et al. 1980. Is photogrammetry of the face reliable? *Plastic and Reconstructive Surgery* 66 3, pp. 346-355.

Farkas, L. G. 1981. *Anthropometry of the Head and Face in Medicine*. New York: Elsevier.

Farkas, L. G. 1994. *Anthropometry of the Head and Face*. New York: Raven Press.

Farkas, L. G. and Deutsch, C. K. 1996. Anthropometric determination of craniofacial morphology. *American Journal of Medical Genetics* 65(1), pp. 1-4.

Farnell, D. J. J. et al. 2016. Multilevel principal component analysis (mPCA) in shape analysis: A feasibility study in medical and dental imaging. *Computer Methods and Programs in Biomedicine* 129, pp. 149-159.

Farnell, D. J. J. et al. 2017. Initial Results of Multilevel Principal Components Analysis of Facial Shape. In: Valdés Hernández, M. and González-Castro, V. eds.

Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science. Vol. 723. Cham: Springer, pp. 1-12.

Farnell, D. J. J. et al. 2019. Multilevel models of age-related changes in facial shape in adolescents. In: Zheng, Y. et al. eds. *Communications in Computer and Information Science*. Vol. 1065. Cham: Springer, pp. 101–113.

Farnell, D. J. J. et al. 2020. Multilevel prinicpal components analysis of three-dimensional facial growth in adolescents. *Computer Methods and Programs in Biomedicine* 188, p. 105272.

Farnell, D. J. J. et al. 2021. An exploration of adolescent facial shape changes with age via multilevel partial least squares regression. *Computer Methods and Programs in Biomedicine* 200, p. 105935.

Ferguson, M. W. 2000. A hole in the head. *Nature Genetics* 24(4), pp. 330-331.

Ferrario, V. F. et al. 1996. Facial three-dimensional morphometry. *American Journal of Orthodontics and Dentofacial Orthopedics* 109(1), pp. 86-93.

Ferrario, V. F. et al. 1997. Three-dimensional study of growth and development of the nose. *The Cleft Palate-Craniofacial Journal* 34(4), pp. 309-317.

Ferrario, V. F. et al. 1999a. Soft-tissue facial morphometry from 6 years to adulthood: a three-dimensional growth study using a new modeling. *Plastic and Reconstructive Surgery* 103(3), pp. 768-778.

Ferrario, V. F. et al. 1999b. Soft tissue facial growth and development as assessed by the three-dimensional computerized mesh diagram analysis. *American Journal of Orthodontics and Dentofacial Orthopedics* 116(2), pp. 215-228.

Ferrario, V. F. et al. 2000. A three-dimensional quantitative analysis of lips in normal young adults. *The Cleft Palate-Craniofacial Journal* 37(1), pp. 48-54.

Ferrario, V. F. et al. 2001. Morphometry of the orbital region: a soft-tissue study from adolescence to mid-adulthood. *Plastic and Reconstructive Surgery* 108(2), pp. 285-292.

Ferrario, V. F. et al. 2003. Growth and aging of facial soft tissues: A computerized three-dimensional mesh diagram analysis. *Clinical Anatomy* 16(5), pp. 420-433.

Ferrario, V. F. et al. 2004. Soft tissue facial morphology in obese adolescents: a three-dimensional noninvasive assessment. *The Angle Orthodontist* 74(1), pp. 37-42.

Ferrario, V. F. et al. 2009. Labial morphology: a 3-dimensional anthropometric study. *Journal of Oral and Maxillofcaial Surgery* 67(9), pp. 1832-1839.

Ferry, Q. et al. 2014. Diagnostically relevant facial gestalt information from ordinary photos. *eLife* 3, p. e02020.

Fisher, R. A. 1938. The statistical utilization of multiple measurements. *Annals of Eugenics* 8(4), p. 376-386.

Flaherty, K. et al. 2016. Understanding craniosynostosis as a growth disorder. *Wiley Interdisciplinary Reviews: Developmental Biology* 5(4), pp. 429-459.

Franco, F. C. M. et al. 2013. Bradycephalic, dolichocephalic and mesocephalic: Is it appropriate to describe the face using skull patterns? *Dental press journal of orthodontics* 18(3), pp. 159-163.

Fraser, A. et al. 2013. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiol*ogy 42(1), pp. 97-110.

Freeman, M. W. and Walford, G. A. 2016. Chapter 41 - Lipoprotein Metabolism and the Treatment of Lipid Disorders. In: Jameson, J.L. et al. eds. *Endocrinology: Adult and Pediatric (Seventh Edition)*. Philadelphia: W.B. Saunders, p. 715-736.e717.

Fried, P. A., O'Connell, C.M. . 1997. A comparison of the effects of prenatal exposure to tobacco, alcohol, cannabis and caffeine on birth size and subsequent growth. *Neurotoxicology and Teratology* 9(2), pp. 79-85.

Galloway, J. et al. 2020. Multilevel Analysis of the Influence of Maternal Smoking and Alcohol Consumption on the Facial Shape of English Adolescents. *Journal of Imaging* 6(5), p. 34.

Geladi, P. and Kowalski, B. R. 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta* 185, pp. 1-17.

Giant Consortium. 2019. *GIANT: Genetic Investigation of Anthropometric Traits*. Available at: https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium [Accessed: 20th April 2021].

Glass, A. R. et al. 1977. Low serum testosterone and sex-hormone-binding-globulin in massively obese men. *The Journal of Clinical Endocrinology & Metabolism* 45(6), pp. 1211-1219.

Goodhue, D. L. et al. 2012. Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly* 36(3), pp. 981-1001.

Gor, T. et al. 2010. Three-dimensional comparison of facial morphology in white populations in Budapest, Hungary, and Houston, Texas. *American Journal of Orthodontics and Dentofacial Orthopedics* 137(3), pp. 424-432.

Gougoutas, A. J. et al. 2007. Hemifacial microsomia: clinical features and pictographic representations of the OMENS classification system. *Plastic and Reconstructive Surgery* 120(7), p. 112e-113e.

Graw, M. et al. 1999. The form of the supraorbital margin as a criterion in identification of sex from the skull: investigations based on modern human skulls. *American Journal of Physical Anthropology* 108(1), pp. 91-96.

Gray, C. D. and Kinnear, P. R. 2012. *IBM SPSS 19 Statistics made simple*. USA: Psychology Press.

Greene, D. J. et al. 2016. Considerations for MRI study design and implementation in pediatric and clinical populations. *Developmental Cognitive Neuroscience* 18, pp. 101-112.

Guttman, L. 1954. Some necessary conditions for common-factor analysis. *Psychometrika* 19(2), pp. 149-161.

Hackshaw, A. et al. 2011. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. *Human Reproduction Update* 17(5), pp. 589-604.

Haibo, H. et al. eds. 2008. *ADASYN: Adaptive synthetic sampling approach for imbalanced learning*. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence),* pp. 1322-1328.

Hannuksela, A. 1981. The effect of moderate and severe atopy on the facial skeleton. *European Journal of Orthodontics* 3(3), pp. 187-193.

Haque, M. M. et al. 2014. Imbalanced class learning in epigenetics. *Journal of Computational Biology* 21(7), pp. 492-507.

Harari, D. et al. 2010. The effect of mouth breathing versus nasal breathing on dentofacial and craniofacial development in orthodontic patients. *Laryngoscope* 120(10), pp. 2089-2093.

Haselhuhn, M. P. et al. 2015. Men's facial width-to-height ratio predicts aggression: a meta-analysis. *PLoS One* 10(4), p. e0122637-e0122637.

Heart UK. 2021. *What do my results mean?* Available at: https://www.heartuk.org.uk/cholesterol/getting-a-cholesterol-test#test-results [Accessed: 22nd April 2021].

Henderson, J. et al. 2008. Household chemicals, persistent wheezing and lung function: effect modification by atopy? *European Respiratory Journal* 31(3), pp. 547-554.

Hennessy, R. J. et al. 2002. 3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebro-craniofacial morphogenesis: initial application to sexual dimorphism. *Biological Psychiatry* 51(6), pp. 507-514.

Hennessy, R. J. et al. 2006. Facial shape and asymmetry by three-dimensional laser surface scanning covary with cognition in a sexually dimorphic manner. *The Journal of Neuropsychiatry and Clinical Neurosciences* 18(1), pp. 73-80.

Hennessy, R. J. et al. 2007. Three-dimensional laser surface imaging and geometric morphometrics resolve frontonasal dysmorphology in schizophrenia. *Biological Psychiatry* 61(10), pp. 1187-1194.

Hennessy, R. J. et al. 2010. Frontonasal dysmorphology in bipolar disorder by 3D laser surface imaging and geometric morphometrics: comparisons with schizophrenia. *Schizophrenia research* 122(1-3), pp. 63-71.

Hochberg, Z. et al. 1992. The distal axis of growth hormone (GH) in nutritional disorders: GH-binding protein, insulin-like growth factor-I (IGF-I), and IGF-I receptors in obesity and anorexia nervosa. *Metabolism* 41(1), pp. 106-112.

Hoogeveen, R. C. et al. 2015. Dose reduction in orthodontic lateral cephalography: dosimetric evaluation of a novel cephalographic thyroid protector (CTP) and anatomical cranial collimation (ACC). *Dentomaxillofacial Radiology* 44(4), p. 20140260.

Hopkin, G. B. et al. 1968. The cranial base as an aetiological factor in malocclusion. *The Angle Orthodontist* 38(3), pp. 250-255.

Hopman, S. M. et al. 2014. Face shape differs in phylogenetically related populations. *European Journal of Human Genetics* 22(11), pp. 1268-1271.

Hoskens, H. et al. 2018. Spatially dense 3D facial heritability and modules of co-heritability in a father-offspring design. *Frontiers in Genetics* 9(554), pp. 2-62.

Howe, J. L. et al. 2019. Prenatal alcohol exposure and facial morphology in a UK cohort. *Drug and Alcohol Dependence* 197, pp. 42-47.

Isaksson, O. G. et al. 1987. Mechanism of the stimulatory effect of growth hormone on longitudinal bone growth. *Endocrine Reviews* 8(4), pp. 426-438.

Jain, A. et al. 2017. Addressing class imbalance problem in medical diagnosis: A genetic algorithm approach. *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC),* pp.1-8.

Jarvis, D. and Burney, P. 1998. ABC of allergies. The epidemiology of allergic disease. *BMJ (Clinical research ed.)* 316(7131), pp. 607-610.

Jiménez-Silva, A. et al. 2021. Craniofacial growth predictors for class II and III malocclusions: A systematic review. *Clinical and Experimental Dental Research* 7(2), pp. 242-262.

Jin, S.-W. et al. 2016. Development and Growth of the Normal Cranial Vault : An Embryologic Review. *Journal of Korean Neurosurgical Society* 59(3), pp. 192-196.

Johnson, C. Y. and Little, J. 2008. Folate intake, markers of folate status and oral clefts: is the evidence converging? *International Journal of Epidemiology* 37(5), pp. 1041-1058.

Jolliffe, I. T. 2002. *Principal component analysis*. 2nd ed. New York: Springer.

Jones, K. L. and Smith, D. W. 1973. Recognition of the fetal alcohol syndrome in early infancy. *Lancet* 302(7836), pp. 999-1001.

Jonsson, H. et al. 2021. Differences between germline genomes of monozygotic twins. *Nature Genetics* 53(1), pp. 27-34.

Källén, K. 1999. Maternal smoking and craniosynostosis. *Teratology* 60(3), pp. 146-150.

Källén, K. 2000. Maternal smoking during pregnancy and infant head circumference at birth. *Early Human Development* 58(3), pp. 197-204.

Kamak, H. et al. 2013. Cranial base features between sagittal skeletal malocclusions in Anatolian Turkish adults: Is there a difference? *Journal of Orthodontic Research* 1(2), pp. 52-56.

Kasai, K. et al. 1995. Relationship between cranial base and maxillofacial morphology. *European Journal of Orthodontics* 17(5), pp. 403-410.

Kau, C. H. et al. 2004. The feasibility of measuring three-dimensional facial morphology in children. *Orthodontics & Craniofacial Research* 7(4), pp. 198-204.

Kau, C. H. et al. 2005. Reliability of measuring facial morphology with a 3-dimensional laser scanning system. *American Journal Orthodontics and Dentofacial Orthopedics* 128(4), pp. 424-430.

Kau, C. H. et al. 2006. Facial templates: a new perspective in three dimensions. *Orthodontics & Craniofacial Research* 9(1), pp. 10-17.

Kau, C. H. et al. 2010. Use of 3-dimensional surface acquisition to study facial morphology in 5 populations. *American Journal of Orthodontics and Dentofacial Orthopedics* 137(4), p. S56.E51-S56.E59.

Kesterke, M. J. et al. 2016. Using the 3D Facial Norms Database to investigate craniofacial sexual dimorphism in healthy children, adolescents, and adults. *Biology of Sex Differences* 7, p. 23.

Kjellberg, H. et al. 2000. Craniofacial morphology, dental occlusion, tooth eruption, and dental maturity in boys of short stature with or without growth hormone deficiency. *European Journal of Oral Sciences* 108(5), pp. 359-367.

Klebanoff, M. A. et al. 1998. Serum cotinine concentration and self-reported smoking during pregnancy. *American Journal of Epidemiology* 148(3), pp. 259-262.

Klingenberg, C.P. 2021. How exactly did the nose get that long? A critical rethinking of the pinocchio effect and how shape changes relate to landmarks. *Evolutionary Biology* 48, pp. 115-127.

Kotu, V. and Deshpande, B. 2015. Chapter 4 - Classification. In: Kotu, V. and Deshpande, B. eds. *Predictive Analytics and Data Mining*. Boston: Morgan Kaufmann, pp. 63-163.

Koudelová, J. et al. 2015. Development of facial sexual dimorphism in children aged between 12 and 15 years: a three-dimensional longitudinal study. *Orthodontics & Craniofacial Research* 18(3), pp. 175-184.

Kozieł, S. et al. 2018. Parental smoking during pregnancy and head shape and size in school children. *Annuals of Human Biol* 45(5), pp. 401-405.

Kvalvik, L. G. et al. 2012. Self-reported smoking status and plasma cotinine concentrations among pregnant women in the Norwegian Mother and Child Cohort Study. *Pediatric Research* 72(1), pp. 101-107.

Lavergne, J. and Petrovic, A. 1983. Discontinuities in occlusal relationship and the regulation of facial growth. A cybernetic view. *European Journal of Orthodontics* 5(4), pp. 269-278.

Lawlor, D. A. et al. 2010. Association between general and central adiposity in childhood, and change in these, with cardiovascular risk factors in adolescence: prospective cohort study. *BMJ (Clinical research ed.)* 341, p. c6224.

Le Stunff, C. et al. 2000. The insulin gene VNTR is associated with fasting insulin levels and development of juvenile obesity. *Nature Genetics* 26(4), pp. 444-446.

Lecron, F. et al. 2012. Multilevel statistical shape models: A new framework for modeling hierarchical structures. *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI),* pp. 1284-1287.

Lee, B. J. and Kim, J. Y. 2014. A Comparison of the Predictive Power of Anthropometric Indices for Hypertension and Hypotension Risk. *PLoS One* 9(1), p. e84897.

Lee, J. M. et al. 2010. Body Mass Index and Timing of Pubertal Initiation in Boys. *Archives of Pediatrics & Adolescent Medicine* 164(2), pp. 139-144.

Lee, P. A. and Witchel, S. F. 1997. The influence of estrogen on growth. *Current Opinion in Pediatrics* 9(4), pp. 431-436.

Leslie, S. et al. 2015. The fine-scale genetic structure of the British population. *Nature* 519(7543), pp. 309-314.

Linder-Aronson, S. 1974. Effects of adenoidectomy on dentition and nasopharynx. *American Journal of Orthodontics* 65(1), pp. 1-15.

Lindley, A. A. et al. 2000. Effect of continuing or stopping smoking during pregnancy on infant birth weight, crown-heel length, head circumference, ponderal index, and brain:body weight ratio. *American Journal of Epidemiology* 152(3), pp. 219-225.

Lindley, A. A., Becker, S., Gray, R.H., Herman, A.A. . 2000. Effect of continuing or stopping smoking during pregnancy on infant birth weight, crown-heel length, head circumference, pinderal index, and brain:body weight ratio. *American Journal of Epidemiology* 152(3), pp. 219-225.

Lipper, E. et al. 1981. Determinants of neurobehavioral outcome in low-birth-weights infants. *Pediatrics* 67(4), pp. 502-505.

Liu, H. et al. 2002. Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery* 6(4), pp. 393-423.

Longo, L. D. 1976. Carbon monoxide: effects on oxygenation of the fetus in utero. *Science* 194(4264), pp. 523-525.

Luck, W. et al. 1985. Extent of nicotine and cotinine transfer to the human fetus, placenta and amniotic fluid of smoking mothers. *Developmental Pharmacology nd Therapeutics* 8(6), pp. 384-395.

Mah, J. and Hatcher, D. 2003. Current status and future needs in craniofacial imaging. *Orthodontics & Craniofacial Research* 6 Suppl 1, pp. 10-16.

Mahadevan, S. and Ali, I. 2016. Is body mass index a good indicator of obesity? *International Journal of Diabetes in Developing Countries* 36(2), pp. 140-142.

Malloy, M. H. et al. 1989. Maternal smoking during pregnancy: no association with congenital malformations in Missouri 1980-83. *American Journal of Public Health* 79(9), pp. 1243-1246.

Manfredi, C. et al. 1997. Heritability of 39 orthodontic cephalometric parameters on MZ, DZ twins and MN-paired singletons. *American Journal of Orthodontics and Dentofacial Orthopedics* 111(1), pp. 44-51.

Martinelli, C. E. et al. 2008. Physiology of the GH-IGF axis. *Arquivos Brasileiros de Endocrinologia & Metabologia* 52(5), pp. 717-725.

Mathur, S. and Sutton, J. 2017. Personalized medicine could transform healthcare. *Biomedical reports* 7(1), pp. 3-5.

Matthews, H. et al. 2016. Spatially dense morphometrics of craniofacial sexual dimorphism in 1-year-olds. *Journal of Anatomy* 229(4), pp. 549-559.

Medawar, P. B. and Fisher, R. A. 1944. The shape of the human being as a function of time. *Proceedings of the Royal Society of London. Series B - Biological Sciences* 132(867), pp. 133-141.

Mehmood, T. et al. 2012. A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems* 118, pp. 62-69.

Mehmood, T. et al. 2020. Comparison of variable selection methods in partial least squares regression. *Journal of Chemometrics* 34(6), p. e3226.

Melsen, B. 1972. Time and mode of closure of the spheno-occipital synchrondrosis determined on human autopsy material. *Acta Anatomica (Basel)* 83(1), pp. 112-118.

Melsen, B. 1974. The cranial base: the postnatal development of the cranial base studied histologically on human autopsy maternal. *Acta Odontologica Scandinavica* 32, pp. 1-12.

Mercier, S. et al. 2011. New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *Journals of Medical Genetics* 48(11), pp. 752-760.

Montavon, G. et al. 2018. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing* 73, pp. 1-15.

Moore, K. L. and Dalley, A. F. 2006. Head. In: *Clinically Orientated Anatomy.* Baltimore: Lippincott Williams & Williams, pp. 886-1038.

Moore, M. K. 2013. Chapter 4 – Sex estimation and assessment. In: *Research Methods in Human Skeletal Biology*, Oxford: Elsevier, pp. 91-116.

Moss, M. L. and Salentijn, L. 1969. The primary role of functional matrices in facial growth. *American Journal of Orthodontics* 55(6), pp. 566-577.

Mossey, P. A. et al. 2009. Cleft lip and palate. *Lancet* 374(9703), pp. 1773-1785.

Muggli, E. et al. 2017. Association between prenatal alcohol exposure and craniofacial shape of children at 12 months of age. *JAMA Pediatrics* 171(8), pp. 771-780.

Mydlová, M. et al. 2015. Sexual dimorphism of facial appearance in ageing human adults: A cross-sectional study. *Forensic Science International* 257, p. 519.e511-519.e519.

Naqvi, S. et al. 2020. Shared heritability of face and brain shape distinct from cognitive traits. *bioRxiv*, p. 2020.2008.2029.269258.

Nie, X. 2005. Cranial base in craniofacial development: developmental features, influence on facial growth, anomaly, and molecular basis. *Acta Odontologica Scandinavica* 63(3), pp. 127-135.

Nunes, L. A. et al. 2018. Geometric morphometrics and face shape characteristics associated with chronic disease in the elderly. *Bioscience Journal* 34(2), pp. 435-446.

Nute, S. J. and Moss, J. P. 2000. Three-dimensional facial growth studied by optical surface scanning. *Journal of Orthodontics* 27(1), pp. 31-38.

O'Brien, K. et al. 2003. Effectiveness of early orthodontic treatment with the Twin-block appliance: a multicenter, randomized, controlled trial. Part 1: Dental and skeletal

effects. *American Journal of Orthodontics and Dentofacial Orthopedics* 124(3), pp. 234-243.

O'Toole, A. J. et al. 1997. Sex Classification is Better with Three-Dimensional Head Structure Than with Image Intensity Information. *Perception* 26(1), pp. 75-84.

Ohrn, K. et al. 2002. Craniofacial morphology in obese adolescents. *Acta Odontologica Scandinavica* 60(4), pp. 193-197.

Ong, K. K. et al. 2009. Infancy weight gain predicts childhood body fat and age at menarche in girls. *The Journal of Clinical Endocrinology & Metabolism* 94(5), pp. 1527-1532.

Oshiro, T. M. et al. 2012. How Many Trees in a Random Forest*?* In: Perner, P. ed. *Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Lecture Notes in Computer Science*, vol 7367. Berlin: Springer.

Papadopoulos, M. A. et al. 2002. Three-dimensional craniofacial reconstruction imaging. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology* 93(4), pp. 382-393.

Pauwels, R. et al. 2012. Effective dose range for dental cone beam computed tomography scanners. *European Journal of Radiology* 81(2), pp. 267-271.

Peng, J. et al. 2005. Craniofacial morphology in Chinese female twins: a semi-longitudinal cephalometric study. *European Journal of Orthodontics* 27(6), pp. 556-561.

Petleshkova, T. D. et al. 2013. Sex-related differences in the lips and mouth area of Buglarians - an investigation by 3D laser scanning. *Scripta Scientifica Medica* 45, pp. 97-100.

Petrovik, A. G. et al. 1981. The final length of the mandible. Is it genetically predetermined? In: Carlson, D. S. *Craniofacial Biology*. Ann Arbor: University of Michigan.

Petrovik, A. G. et al. 1975. Control processes in the postnatal growth of the condylar cartilage of the mandible. In: McNamara, J.A. ed. *Determinants of mandibular form and growth. Monograph No 4: Craniofacial Growth Series, Centre for Human Growth and Development*. Ann Arbor: University of Michigan.

Pirinen, S. et al. 1994. Craniofacial features in patients with deficient and excessive growth hormone. *Journal of Craniofacial Genetics and Developmental Biology* 14(3), pp. 144-152.

Plymate, S. R. et al. 1988. Inhibition of sex hormone-binding globulin production in the human hepatoma (Hep G2) cell line by insulin and prolactin. *The Journal of Clinical Endocrinology & Metabolism* 67(3), pp. 460-464.

Polat, O. O. and Kaya, B. 2007. Changes in cranial base morphology in different malocclusions. *Orthodontics & Craniofacial Research* 10(4), pp. 216-221.

Poswillo, D. 1973. The pathogenesis of the first and second branchial arch syndrome. *Oral Surgery, Oral Medicine, Oral Pathology* 35(3), pp. 302-328.

Pound, N. et al. 2014. Facial fluctuating asymmetry is not associated with childhood ill-health in a large British cohort study. *Proceedings of the Royal Society B: Biological Sciences* 281(1792), p. 20141639.

Prince, C. et al. 2019. Investigating the impact of cigarette smoking behaviours on DNA methylation patterns in adolescence. *Human Molecular Genetics* 28(1), pp. 155-165.

Qiu, H. et al. 2017. Influence of insulin on growth hormone secretion, level and growth hormone signalling. *Sheng Li Xue Bao* 69(5), pp. 541-556.

Quinonez, S. C. and Innis, J. W. 2014. Human HOX gene disorders. *Molecular Genetics and Metabolism* 111(1), pp. 4-15.

Rabey, G. 1971. Craniofacial morphanalysis. *Proceedings of the Royal Society of Medicine* 64(2), pp. 103-111.

Rahman, M. S. et al. 2021. Role of Insulin in Health and Disease: An Update. *International Journal of Molecular Sciences* 22(12), p. 6403.

Rantala, M. J. et al. 2013. Adiposity, compared with masculinity, serves as a more valid cue to immunocompetence in human mate choice. *Proceedings of the Royal Society B: Biological Sciences* 280(1751), p. 20122495.

Reddy, M. S. et al. 2019. Embracing Personalized Medicine in Dentistry. *Journal of Pharmacy & Bioallied Sciences* 11(Suppl 2), pp. S92-S96.

Reddy, G. T. et al. 2020. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access* 8, pp. 54776-54788.

Reed, M. J. et al. 1987. Dietary lipids: an additional regulator of plasma levels of sex hormone binding globulin. *The Journal of Clinical Endocrinology & Metabolism* 64(5), pp. 1083-1085.

Rich, S. S. et al. 2005. A Genome Scan for Fasting Insulin and Fasting Glucose Identifies a Quantitative Trait Locus on Chromosome 17p. *Diabetes* 54(1), p. 290.

Richmond, R. C. et al. 2015. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Human Molecular Genetics* 24(8), pp. 2201-2217.

Richmond, S. et al. 2018. Facial genetics: A brief overview. *Frontiers in Genetics* 9(462), pp.1-21.

Richmond, S. et al. 2020. Exploring the midline soft tissue surface changes from 12 to 15 years of age in three distinct country population cohorts. *European Journal of Orthodontics* 42(5), pp. 517-524.

Riedel, M. et al. 1995. Pulsatile growth hormone secretion in normal-weight and obese men: differential metabolic regulation during energy restriction. *Metabolism* 44(5), pp. 605-610.

Riley, E. P. et al. 2011. Fetal alcohol spectrum disorders: an overview. *Neuropsychology review* 21(2), pp. 73-80.

Ringnér, M. 2008. What is principal component analysis? *Nature Biotechnology* 26(3), pp. 303-304.

Romero-Corral, A. et al. 2010. Interactions between obesity and obstructive sleep apnea: implications for treatment. *Chest* 137(3), pp. 711-719.

Rönning, O. and Koski, K. 1969. The effect of the articular disc on the growth of condylar cartilage transplants. *Report of the Congress. European Orthodontic Society*, pp. 99-108.

Roosenboom, J. et al. 2018. SNPs associated with testosterone levels influence human facial morphology. *Frontiers in Genetics* 9(497), pp. 1-10.

Ryöppy, S. 1965. Transplantation of epiphyseal cartilage and cranial suture. Experimental studies on the preservation of the growth capacity in growing bone grafts. *Acta Orthopopaedica Scandcandinavica Supplementum*, Suppl 82, pp. 81-106.

Sadeghianrizi, A. et al. 2005. Craniofacial development in obese adolescents. *European Journal of Orthodontics* 27(6), pp. 550-555.

Sadler, T. W. 2012. Head and Neck. In: Sadler, T. W. ed. *Langman's Medical Embryology*. 12th ed. Baltimore: Lippincott Williams & Williams, pp. 260-286.

SAGE. 2011. Encyclopedia of Measurement and Statistics. In: Salkind, N.J. ed. *Encyclopedia of Measurement and Statistics*. California: Sage Publications, pp. 741-744.

Sassouni, V. et al. 1985. The influence of perennial allergic rhinitis on facial type and a pilot study of the effect of allergy management on facial growth patterns. *Annals of Allergy* 54(6), pp. 493-497.

Savoye, I. et al. 1998. A genetic study of anteroposterior and vertical facial proportions using model-fitting. *The Angle Orthodontist* 68(5), pp. 467-470.

Scheuer, L. 2002. Application of osteology to forensic medicine. *Clinical Anatomy* 15(4), pp. 297-312.

Scheuer, L. and Black, S. 2007a. *Forensic Human Identification: An Introduction*. Florida: CRC Press.

Scheuer, L. and Black, S. 2007b. Osteology. In: Thompson, T. and Black, S. eds. *Forensic Human Identification: An Introduction*. Florida: CRC Press.

Schulhof, R. J. et al. 1977. Prediction of abnormal growth in class III malocclusions. *American Journal of Orthodontics* 71(4), pp. 421-430.

Scott, J. H. 1953. The cartilage of the nasal septum. *British Dental Journal* 95, pp. 37-43.

Scott, J. H. 1954. The growth of the human face. *Proceedings of the Royal Societyof Medicine* 47, pp. 91-100.

Scott, J. H. 1956. Growth at the facial sutures. *American Journal of Orthodontics* 42, pp. 381-387.

Selby, C. 1990. Sex hormone binding globulin: origin, function and clinical significance. *Annals of Clinical Biochemistry* 27 (Pt 6), pp. 532-541.

Sha, T. et al. 2011. Feature level analysis for 3D facial expression recognition. *Neurocomputing* 74(12), pp. 2135-2141.

Shaffer, J. R. et al. 2016. Genome-wide association study reveals multiple loci influencing normal human facial morphology. *PLoS Genetics* 12(8), p. e1006149.

Shaw, G. M. et al. 2002. Maternal periconceptional vitamins: interactions with selected factors and congenital anomalies? *Epidemiology* 13(6), pp. 625-630.

Shi, M. et al. 2007. Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. *American journal of human genetics* 80(1), pp. 76-90.

Shipton, D., Tappin, D.M., Vadiveloo, T., Crosslet, J.A., Aitken, D.A., Chalmers, J. . 2009. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. *British Medical Journal,* 339, p. b4347.

Shrimpton, S. et al. 2014. A spatially-dense regression study of facial form and tissue depth: towards an interactive tool for craniofacial reconstruction. *Forensic Science International* 234, pp. 103-110.

Silva, M. E. et al. 1992. Effects of testosterone on growth hormone secretion and somatomedin-C generation in prepubertal growth hormone deficient male patients. *Brazilian Journal of Medical and Biological Research* 25(11), pp. 1117-1126.

Simmons, D. 2008. Epigenetic influences and disease. *Nature Education* 1(1), p. 6.

Simpson, E. H. 1951. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 13(2), pp. 238-241.

Smith, F. et al. 2013. The effect of hypoxia on facial shape variation and disease phenotypes in chicken embryos. *Disease Models & Mechanisms* 6(4), pp. 915-924.

Snyders, J., et al. 2014. Development and comparison of non-rigid surface registration algorithms and extensions. Report no. KUL/ESAT/PSI/1401. PSI: KU Leuven.

Socol, M. L., Manning, F.A., Muata, Y., Druzin, M.L. . 1982. Maternal smoking causes fetal hypoxia: experimental evidence. *American Journal of Obsetrics and Gynecology* 142(2), pp. 214-218.

Soliman, A. et al. 2014. Advances in pubertal growth and factors influencing it: Can we increase pubertal growth? *Indian journal of endocrinology and metabolism* 18(Suppl 1), pp. S53-S62.

Solow, B. and Houston, W. J. 1988. Mandibular rotations: concepts and terminology. *European Journal of Orthodontics* 10(3), pp. 177-179.

St-Jacques, B. et al. 1999. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes & Development* 13(16), pp. 2072-2086.

Stahle, L. and Wold, S. 1990. Multivariate analysis of variance (MANOVA). *Chemometrics and Intelligent Laboratory Systems* 9(2), pp. 127-141.

Stegmann, M. B. and Gomez, D. D. 2002. *A brief introduction to statistical shape analysis*. Available at: http://graphics.stanford.edu/courses/cs164-09-spring/Handouts/paper_shape_spaces_imm403.pdf [Accessed: 30th Dec 2020].

Stella, O. 2019. Discriminant analysis: An analysis of its predictship function. *Journal of Eduction and Practice* 10(5), pp. 50-57.

Stiby, A. I. et al. 2013. Association of maternal smoking with child cotinine levels. *Nicotine & Tobacco Research* 15(12), pp. 2029-2036.

Sunhem, W. and Pasupa, K. 2016. An approach to face shape classification for hairstyle recommendation. *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)* 2016, pp. 390-394.

Tabachnick, B. G. and Fidell, L. S. 1996. *Using multivariate statistics*. 3rd ed. New York: Harper Collins.

Tan, D. W. et al. 2017. Hypermasculinised facial morphology in boys and girls with Autism Spectrum Disorder and its association with symptomatology. *Scientific Reports* 7(1), p. 9348.

Taylor, A. E. et al. 2014. Partner smoking and maternal cotinine during pregnancy: implications for negative control methods. *Drug and Alcohol Dependence* 139(100), pp. 159-163.

Thornhill, R. and Gangestad, S. W. 2006. Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior* 27(2), pp. 131-144.

Timmerman, M. E. 2006. Multilevel component analysis. *British Journal of Mathematical and Statistical Psychology* 59(Pt 2), pp. 301-320.

Toma, A. M. 2014. *Characterization of normal facial featues and their association with genes.* PhD Thesis, Cardiff University.

Toma, A. M. et al. 2009. Reproducibility of facial soft tissue landmarks on 3D laser-scanned facial images. *Orthodontics & Craniofacial Research* 12(1), pp. 33-42.

Toma, A. M. et al. 2012. The assessment of facial variation in 4747 British school children. *European Journal of Orthodontics* 34(6), pp. 655-664.

Trask, G. M. et al. 1987. The effects of perennial allergic rhinitis on dental and skeletal development: a comparison of sibling pairs. *American Journal of Orthodontics and Dentofacial Orthopedics* 92(4), pp. 286-293.

Van Dongen, S. et al. 2020. Lack of correlation between facial sexual dimorphism, fluctuating asymmetry and self-perceived attractiveness in men and women. *Symmetry* 12(2), p. 236.

Van Erum, R. et al. 1998. Craniofacial growth and dental maturation in short children born small for gestational age: effect of growth hormone treatment. Own observations and review of the literature. *Hormone Research* 50(3), pp. 141-146.

Velemínská, J. et al. 2012. Surface facial modelling and allometry in relation to sexual dimorphism. *HOMO* 63(2), pp. 81-93.

Vignolo, M. et al. 1988. Growth and development in simple obesity. *European Journal of Pediatrics* 147(3), pp. 242-244.

Vuolo, M. and Staff, J. 2013. Parent and child cigarette use: a longitudinal, multigenerational study. *Pediatrics* 132(3), p. e568-577.

Wang, X. et al. 1997. Maternal smoking during pregnancy, urine cotinine concentrations, and birth outcomes. A prospective cohort study. *International Journal of Epidemiology* 26(5), pp. 978-988.

Wang, Y. 2002. Is obesity associated with early sexual maturation? A comparison of the association in American boys versus girls. *Pediatrics* 110(5), pp. 903-910.

Wang, Y. et al. 2016. Auto-encoder based dimensionality reduction. *Neurocomputing* 184, pp. 232-242.

Ward, R. E. 1994. Craniofacial anthropometry in clinical genetics. In: Farkas, L.G. ed. *Anthropometry of the Head and Face*. New York Raven Press, pp. 119-124.

Weinberg, S. M. et al. 2013. Heritability of Face Shape in Twins: A Preliminary Study using 3D Stereophotogrammetry and Geometric Morphometrics. *Dent 3000* 1(1), p. 14.

Weinmann, J. P. and Sicher, H. 1947. *Bone and bones: fundamentals of bone biology*. St Louis: CV Mosby.

Weise, M. et al. 2001. Effects of estrogen on growth plate senescence and epiphyseal fusion. *Proceeding of the National Academy of Sciences USA* 98(12), pp. 6871-6876.

Werler, M. M. et al. 2009. Hemifacial microsomia: from gestation to childhood. *Journal of Craniofacial Surgery* 20 Suppl 1(Suppl 1), pp. 664-669.

Whaites, E. 2002a. Cephalometric radiography. In: Whaites, E. ed. *Essentials of Dental Radiography and Radiology*. 3rd ed. Harcourt: Churchill Livingstone, pp. 145-152.

Whaites, E. 2002b. Radiation protection. In: Whaites, E. ed. *Essentials of dental radiography and radiology*. 3rd ed. Edinburgh: Churchill Livingstone, pp. 53-65.

White, J. D. et al. 2019. MeshMonk: Open-source large-scale intensive 3D phenotyping. *Scientific Reports* 9(1), p. 6085.

White, J. D. et al. 2020. Sources of variation in the 3dMDface and Vectra H1 3D facial imaging systems. *Scientific Reports* 10(1), p. 4443.

White, J. D. et al. 2021. Insights into the genetic architecture of the human face. *Nature Genetics* 53(1), pp. 45-53.

WHO Expert Consultation. 2004. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 363(9403), pp. 157-163.

Wilhelm, B. M. et al. 2001. A comparison of cranial base growth in class I and class II skeletal patterns. *American Journal of Orthodontics and Dentofacial Orthopedics* 119(4), pp. 401-405.

Wold, H. (1966) Estimation of principal components and related models by iterative least squares. In: Krishnaiaah, P.R. ed. *Multivariate Analysis*. New York: Academic Press, pp. 391-420.

Wold, S. et al. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58(2), pp. 109-130.

Wolfram-Gabel, R. and Kahn, J. L. 2002. Adipose body of the orbit. *Clinical Anatomy* 15(3), pp. 186-192.

Woods, S. E. and Raju, U. 2001. Maternal smoking and the risk of congenital birth defects: a cohort study. *The Journal of the American Board of Family Practice* 14(5), pp. 330-334.

Woolcott, O. O. and Bergman, R. N. 2018. Relative fat mass (RFM) as a new estimator of whole-body fat percentage ─ A cross-sectional study in American adult individuals. *Scientific Reports* 8(1), p. 10980.

Xia, Y. 2020. Correlation and association analyses in microbiome study integrating multiomics in health and disease. In: Sun, J. ed. *The Microbiome in Health and Disease*. Oxford: Academic Press, pp. 312-346.

Xu, Y. and Goodacre, R. 2018. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of analysis and testing* 2(3), pp. 249-262.

Yeniay, Ö. and Göktas, A. 2002. A comparison of partial least squares regression with other prediction methods. *Hacettepe Journal of Mathematics and Statistics*, 31, pp. 99-111.

Yong, A. G. and Pearce, S. 2013. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology* 9(2), pp. 79-94.

Zhou, H. et al. 2018. t-Distributed Stochastic Neighbor Embedding Method with the Least Information Loss for Macromolecular Simulations. *Journal of Chemical Theory and Computation* 14(11), pp. 5499-5510.

Zhurov, A. et al. 2005. Computer methods for measuring 3D facial morphology. In: Middleton, J. et al. eds. *Computer Methods in Biomechanics and Biomedical Engineering-5*. Cardiff: First Numerics Ltd, pp. 2-7.

Zhurov, A. I. et al. 2010. Averaging facial images. In: Kau, C.H. and Richmond, S. eds. *Three-dimensional imaging for orthodontics and maxillofacial surgery*. London: Wiley-Blackwell, pp. 126-214.

11 APPENDIX 1: STANDARDISATION PRIOR TO MPCA

11.1 INTRODUCTION

Standardisation involves:

$$Standardised\ landmark = \frac{Orignal\ landmark - mean\ of\ landmark}{Standard\ deviation\ of\ landmark}$$

The influence of standardising the landmark data prior to mPCA was investigated here.

11.2 METHODOLOGY

- Geographical location: Croatian, English, Welsh and Finnish males and females as detailed in the methodology section
- Landmarks: 21 facial landmarks in three dimensions (63 total landmarks)
- Variables: Geographical location and sex
- Analysis 1: Landmarks not standardised first
- Analysis 2: Landmarks standardised first
- mPCA:
    - 3-level non-nested mPCA
    - Median averaging of the covariance matrices
    - Eigenvalues retained: level 1 (geographical location) – 3PCs, level 2 (sex) – 1PC, level 3 (average within-group variation) – 20PCs.

11.3 RESULTS

Following standardisation, the landmarks are reconfigured into a sphere (Figure 74). Standardising the landmarks reduces the magnitude of the within-group variation. This increases the relative importance of population and sex (Figure 75, Table 25). It is also challenging to interpret the clinical differences that each PC explains. This cannot be visualised in the same manner as the current methodology as very little difference is seen between the groups (Figure 76). However, the standardised component scores are separated more clearly by group when standardisation is also carried out prior to mPCA (Figure 77).

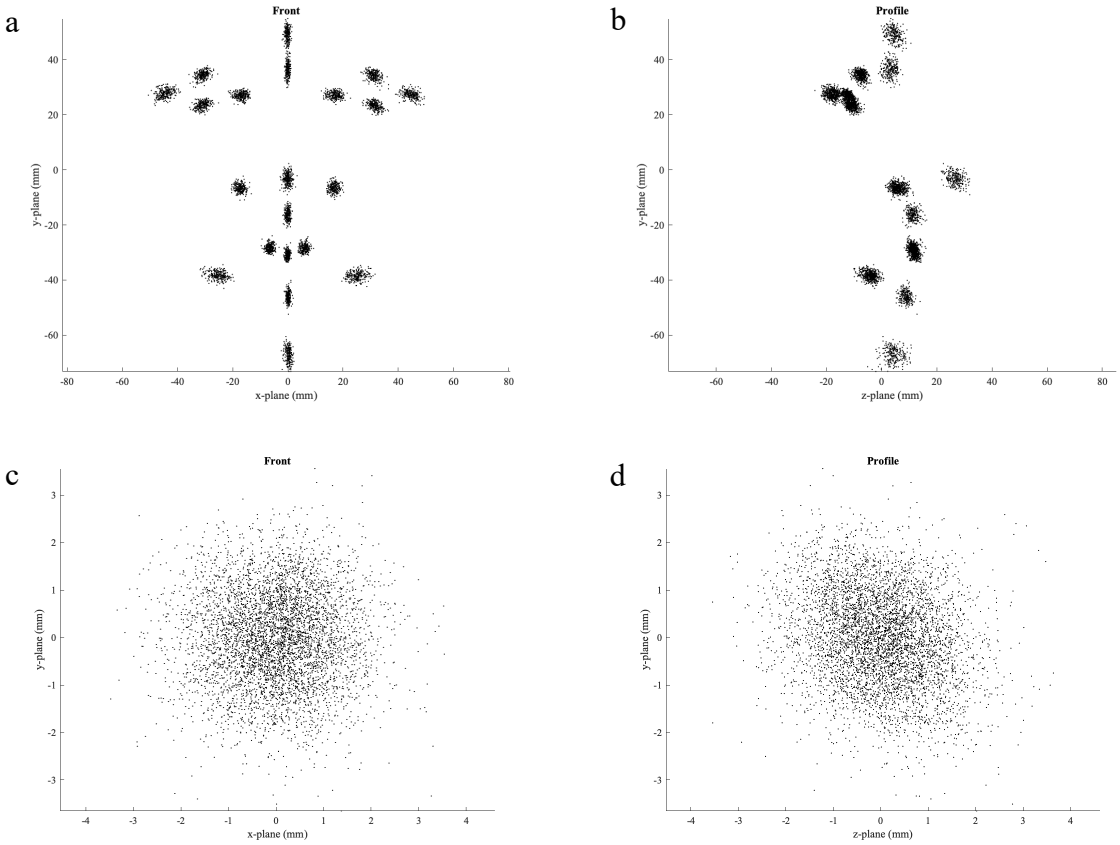


Figure 74: Landmark data without standardisation (a,b) and after standardisation (c,d).

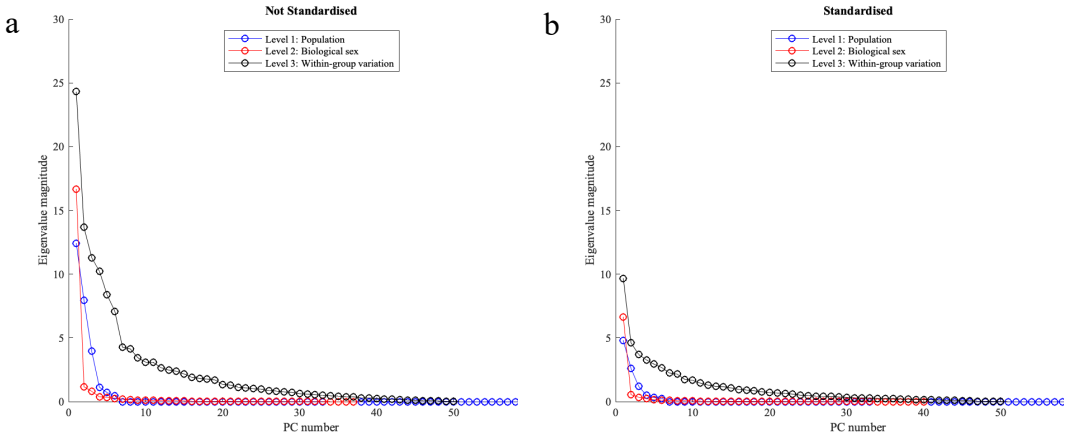


Figure 75: Eigenvalue plot of eigenvalue magnitudes without standardisation (a) and with standardisation (b). The overall pattern is the same, but with standardisation, the eigenvalue magnitudes are reduced. This is to be expected following standardisation.

Table 25: Percentage variation explained by each level when the landmarks are not standardised compared to standardising prior to mPCA.

	% Total variation explained by the retained PCs	
	Without standardisation	With standardisation
Geographical location	14.59	12.55
Sex	9.98	9.69
Within-group variation	66.67	65.84

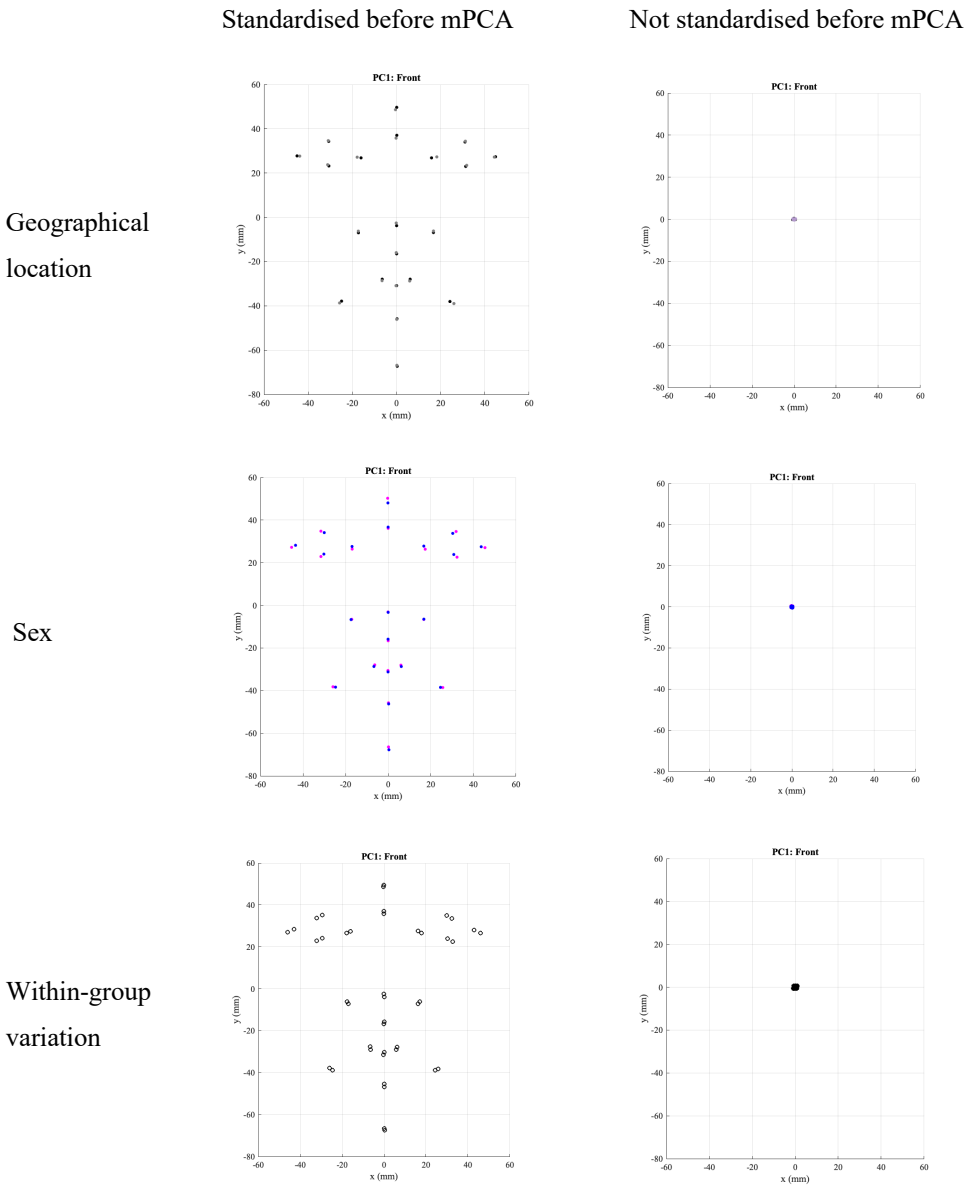


Figure 76: The clinical differences using this method when the landmarks are not standardised before mPCA compared to standardising. PC1 only is displayed to demonstrate. This method is only appropriate for non-standardised landmarks and is easy to understand.

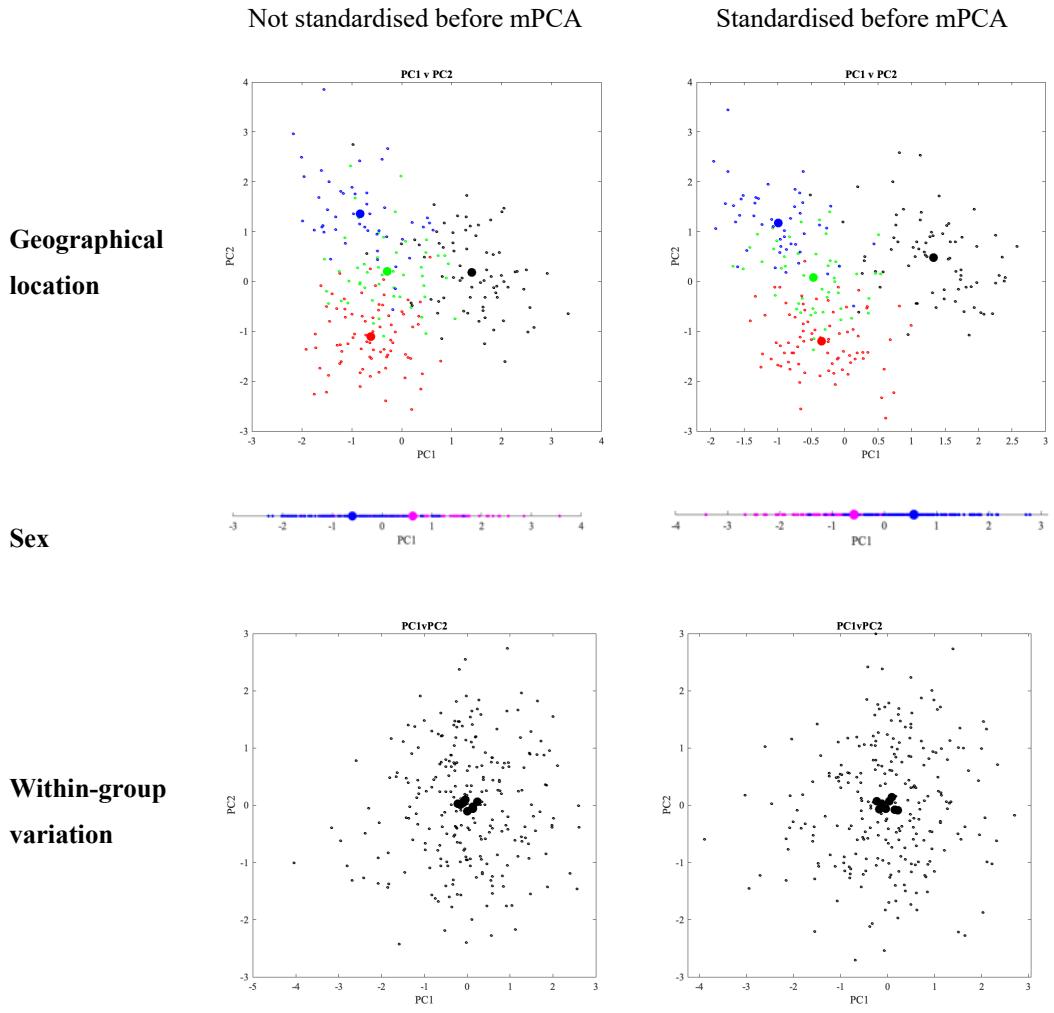


Figure 77: Standardised component score plots (PC1 v PC2, where applicable) for the without standardising of landmark analysis and with standardising of landmarks analysis.

11.4 Conclusion

The standardised component scores are more clearly separated when standardisation is also carried out prior to mPCA. However, the difficultly in interpreting the clinical meaning of the PCs and the exaggerated importance of levels one (geographical location) and two (sex) reduces the usefulness of this step. On balance, it was decided not to standardise the landmarks prior to mPCA for the analyses in the rest of this thesis.

12 APPENDIX 2: SUPPLEMENTARY TABLES

Table 26: Raw landmarks grouped by geographical location. Mean of each landmark +/- standard deviation.

Red= difference >2mm between at least one of the populations, Blue = difference >1mm but < 2mm between at least one of the populations

Glabella (g) Nasion (n); Endocanthion left (enL); Endocanthion right (enR); Exocanthion left (exL); Exocanthion right (exR); Palpebrale superius left (psL); Palpebrale superius right (psR); Palpebrale inferius left (piL); Palpebrale inferius right (piR); Pronasale (prn); Subnasale (sn); Alare left (alL); Alare right (alR); Labiale superius (ls); Crista philtri left (cphL); Crista philtri right (cphR); Labiale inferius (li); Cheilion left (chL); Cheilion right (chR); Pogonion (p).

	Croatian		English		Welsh		Finnish	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
gX	0.204	0.505	-0.357	0.547	-0.333	0.471	-0.361	0.548
gY	49.890	2.039	48.264	1.950	49.712	2.407	48.851	2.190
gZ	4.903	1.771	3.920	1.859	4.032	1.857	5.110	1.998
nX	0.192	0.429	-0.104	0.486	-0.072	0.442	-0.215	0.439
nY	36.855	2.181	36.673	2.143	37.602	2.024	34.046	2.384
nZ	3.579	1.589	2.840	1.707	2.927	1.537	3.166	1.628
enLX	15.493	1.458	17.362	1.443	17.617	1.408	18.440	1.280
enLY	26.736	1.040	26.985	1.076	27.482	1.160	27.316	0.922
enLZ	-11.682	1.129	-12.604	1.327	-12.668	1.112	-12.652	1.079
enRX	-15.881	1.247	-16.775	1.320	-17.519	1.257	-18.357	1.656
enRY	26.827	1.236	26.596	1.034	27.448	1.140	27.492	0.860
enRZ	-11.719	1.140	-11.767	1.229	-12.169	1.156	-12.442	1.201

exLX	45.087	1.993	44.488	1.931	44.164	1.989	44.534	1.875
exLY	27.439	1.419	26.845	1.367	27.344	1.266	27.704	1.124
exLZ	-18.854	1.507	-17.351	1.451	-17.229	1.398	-17.815	1.361
exRX	-45.467	2.059	-44.181	1.991	-43.637	1.756	-44.321	1.873
exRY	27.901	1.521	27.143	1.439	27.499	1.326	28.525	1.249
exRZ	-19.479	1.485	-17.346	1.418	-17.134	1.379	-17.692	1.542
psLX	31.115	1.725	31.297	1.551	30.687	1.655	31.356	1.525
psLY	33.842	1.516	34.667	1.282	34.393	1.171	33.786	1.594
psLZ	-8.071	1.244	-8.021	1.400	-7.145	1.247	-7.272	1.248
psRX	-30.804	1.652	-30.926	1.563	-30.578	1.600	-31.022	1.500
psRY	34.179	1.473	34.747	1.153	34.591	1.239	34.286	1.375
psRZ	-7.926	1.178	-7.798	1.413	-7.035	1.060	-7.090	1.227
piLX	31.355	1.555	31.758	1.566	31.399	1.684	31.868	1.434
piLY	22.959	1.255	23.164	1.332	23.140	1.203	23.856	1.022
piLZ	-11.081	1.237	-10.118	1.469	-11.288	1.320	-11.822	1.198
piRX	-30.642	1.808	-31.112	1.531	-30.723	1.574	-31.282	1.604
piRY	23.165	1.323	23.441	1.146	23.408	1.276	24.082	1.130
piRZ	-10.979	1.222	-9.959	1.495	-11.134	1.361	-11.773	1.140
prnX	0.073	1.166	-0.162	0.905	0.058	0.995	-0.124	0.804
prnY	-4.003	2.002	-2.463	1.776	-3.457	2.059	-2.846	1.869
prnZ	28.078	2.030	26.314	1.797	26.557	1.647	25.593	2.044
snX	0.085	0.629	-0.152	0.614	0.029	0.646	-0.101	0.531
snY	-16.507	1.764	-16.136	1.898	-16.580	1.974	-15.623	1.505
snZ	11.724	1.553	11.083	1.444	11.443	1.403	10.906	1.472
alLX	16.684	1.337	16.768	1.205	16.885	1.188	16.640	1.174
alLY	-7.003	1.347	-6.231	1.389	-6.610	1.621	-6.274	1.343
alLZ	6.934	1.849	5.236	1.407	4.982	1.335	5.782	1.488
alRX	-17.184	1.376	-17.459	1.254	-17.364	1.345	-17.076	1.102

alRY	-7.116	1.236	-6.193	1.234	-6.796	1.355	-6.364	1.319
alRZ	7.369	1.910	5.534	1.469	5.153	1.388	6.077	1.664
lsX	0.131	0.523	-0.205	0.491	-0.256	0.549	-0.060	0.460
lsY	-30.905	1.144	-30.327	1.358	-31.377	1.007	-31.301	1.306
lsZ	11.651	1.034	12.303	1.151	12.230	0.913	12.927	1.157
liX	0.266	0.555	0.103	0.388	0.155	0.523	0.171	0.523
liY	-45.700	1.680	-46.430	1.826	-46.154	1.932	-45.491	1.813
liZ	8.712	1.417	8.846	1.382	8.793	1.583	9.436	1.137
cphLX	6.190	0.999	6.272	1.148	5.826	0.926	5.807	0.841
cphLY	-27.793	1.146	-28.399	1.321	-28.792	0.997	-28.740	1.332
cphLZ	10.794	0.875	11.321	1.100	11.292	0.858	11.906	0.949
cphRX	-6.401	1.095	-6.633	1.002	-6.626	1.033	-6.366	0.738
cphRY	-27.835	1.182	-28.289	1.226	-28.611	0.935	-28.689	1.329
cphRZ	10.972	0.917	11.391	1.056	11.243	0.818	11.846	0.996
chLX	23.710	1.944	26.833	1.976	24.797	1.895	24.753	2.029
chLY	-37.897	1.246	-38.339	1.308	-39.059	1.411	-39.158	1.287
chLZ	-5.093	2.083	-3.754	1.538	-4.077	1.570	-4.016	1.892
chRX	-24.550	1.770	-27.225	2.253	-24.475	1.880	-24.382	2.268
chRY	-37.786	1.301	-38.060	1.310	-38.612	1.441	-39.067	1.255
chRZ	-5.124	2.004	-3.604	1.502	-3.726	1.680	-3.773	1.904
pgX	0.344	0.681	0.411	0.763	-0.036	0.836	0.098	0.753
pgY	-67.248	2.190	-67.658	2.533	-66.571	2.795	-66.393	2.033
pgZ	5.293	2.194	3.533	2.775	4.954	2.700	3.599	2.547

Table 27: Raw landmarks grouped by sex. Mean of each landmark +/- standard deviation.

Blue = >1mm but <2mm difference between the sexes

Glabella (g) Nasion (n); Endocanthion left (enL); Endocanthion right (enR); Exocanthion left (exL); Exocanthion right (exR); Palpebrale superius left (psL); Palpebrale superius right (psR); Palpebrale inferius left (piL); Palpebrale inferius right (piR); Pronasale (prn); Subnasale (sn); Alare left (alL); Alare right (alR); Labiale superius (ls); Crista philtri left (cphL); Crista philtri right (cphR); Labiale inferius (li); Cheilion left (chL); Cheilion right (chR); Pogonion (p).

	Female		**Male**	
	Mean	**SD**	**Mean**	**SD**
gX	-0.230	0.631	-0.148	0.516
gY	49.920	2.047	48.382	2.124
gZ	3.709	1.712	5.184	1.836
nX	-0.038	0.499	-0.026	0.451
nY	36.218	2.452	36.612	2.487
nZ	2.530	1.535	3.727	1.532
enLX	17.229	1.762	16.913	1.776
enLY	26.549	0.916	27.587	0.996
enLZ	-11.794	1.166	-12.904	1.087
enRX	-16.989	1.565	-16.933	1.667
enRY	26.535	1.055	27.462	1.052
enRZ	-11.397	1.085	-12.512	1.067
exLX	45.214	1.818	44.014	1.933
exLY	27.126	1.309	27.434	1.380
exLZ	-17.559	1.636	-18.144	1.479
exRX	-45.169	1.737	-43.798	2.111
exRY	27.369	1.432	28.018	1.473

exRZ	-17.754	1.814	-18.228	1.644
psLX	31.650	1.513	30.627	1.583
psLY	34.518	1.287	33.898	1.519
psLZ	-7.735	1.335	-7.702	1.376
psRX	-31.377	1.418	-30.313	1.564
psRY	34.742	1.212	34.188	1.381
psRZ	-7.491	1.322	-7.605	1.278
piLX	32.085	1.430	31.104	1.552
piLY	22.758	1.249	23.690	1.099
piLZ	-10.215	1.333	-11.681	1.189
piRX	-31.409	1.468	-30.458	1.685
piRY	23.073	1.270	23.866	1.116
piRZ	-10.095	1.335	-11.561	1.230
prnX	-0.086	0.900	0.001	1.071
prnY	-3.038	1.946	-3.331	2.068
prnZ	26.165	1.981	27.309	2.043
snX	-0.025	0.596	-0.048	0.636
snY	-16.482	1.831	-15.998	1.803
snZ	10.909	1.322	11.701	1.563
alLX	16.578	1.207	16.904	1.239
alLY	-6.646	1.341	-6.440	1.543
alLZ	5.605	1.741	5.962	1.708
alRX	-17.026	1.258	-17.541	1.264
alRY	-6.610	1.363	-6.624	1.294
alRZ	5.937	1.788	6.254	1.886
lsX	-0.094	0.544	-0.085	0.513
lsY	-30.631	1.229	-31.145	1.297
lsZ	12.208	1.092	12.222	1.217

liX	0.147	0.541	0.200	0.448
liY	-45.867	1.657	-46.097	1.988
liZ	9.029	1.279	8.789	1.522
cphLX	5.990	0.979	6.149	1.063
cphLY	-28.153	1.261	-28.571	1.251
cphLZ	11.200	1.026	11.340	1.032
cphRX	-6.383	0.959	-6.641	1.017
cphRY	-28.051	1.181	-28.535	1.216
cphRZ	11.267	0.981	11.380	1.021
chLX	25.478	2.431	24.762	2.149
chLY	-38.526	1.291	-38.491	1.497
chLZ	-4.713	1.905	-3.819	1.710
chRX	-25.719	2.491	-24.994	2.276
chRY	-38.256	1.373	-38.305	1.431
chRZ	-4.489	1.817	-3.732	1.875
pgX	0.174	0.852	0.309	0.677
pgY	-66.549	2.194	-67.600	2.571
pgZ	4.683	2.575	4.019	2.725

Table 28: Raw landmarks grouped by geographical location and sex. Mean of each landmark +/- standard deviation.

Red = >2SD, blue = <1SD

Glabella (g) Nasion (n); Endocanthion left (enL); Endocanthion right (enR); Exocanthion left (exL); Exocanthion right (exR); Palpebrale superius left (psL); Palpebrale superius right (psR); Palpebrale inferius left (piL); Palpebrale inferius right (piR); Pronasale (prn); Subnasale (sn); Alare left (alL); Alare right (alR); Labiale superius (ls); Crista philtri left (cphL); Crista philtri right (cphR); Labiale inferius (li); Cheilion left (chL); Cheilion right (chR); Pogonion (p).

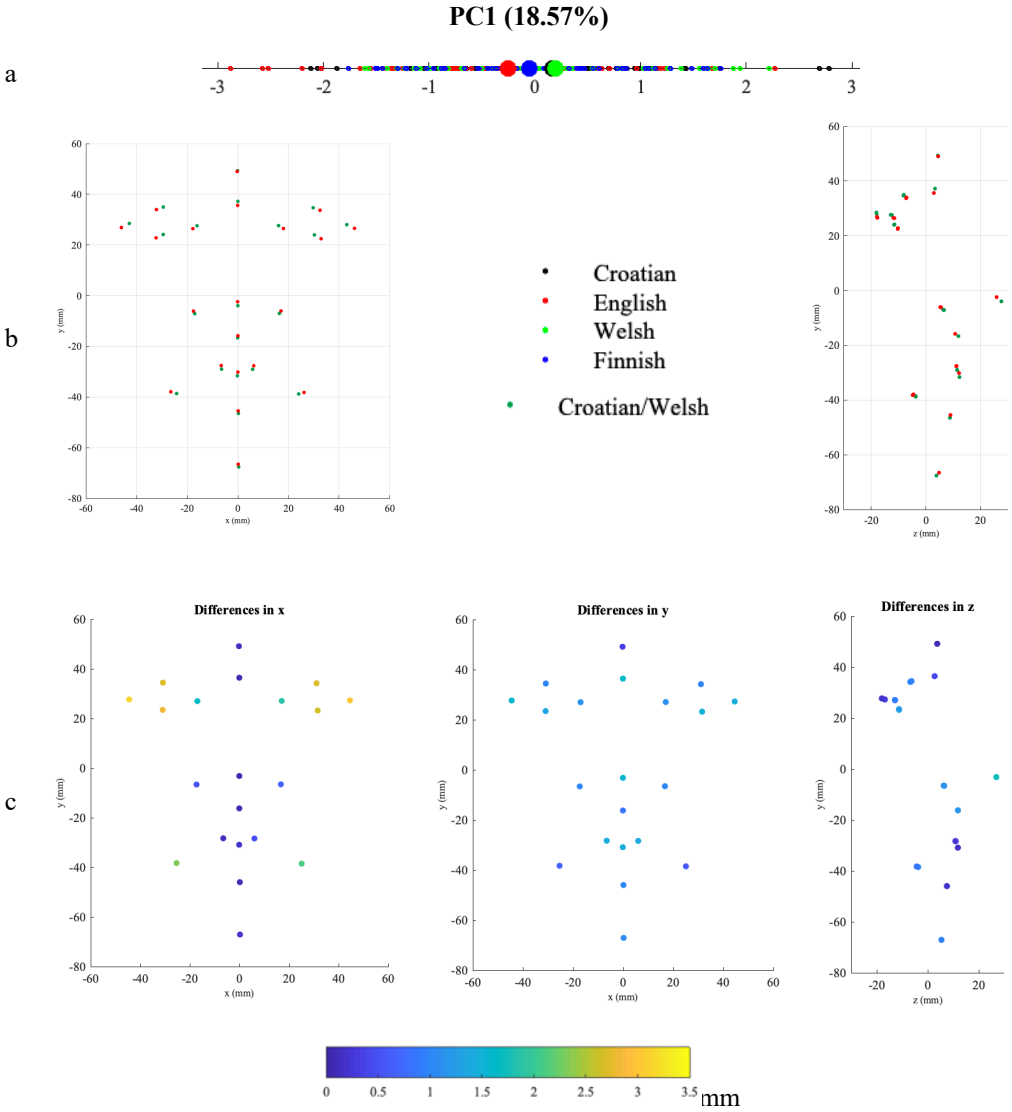
	Croatian Female		Croatian Male		English Female		English Male		Welsh Female		Welsh Male		Finnish Female		Finnish Male	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
gX	0.170	0.528	0.241	0.484	-0.437	0.607	-0.279	0.476	-0.292	0.541	-0.369	0.409	-0.477	0.625	-0.248	0.448
gY	50.618	1.898	49.099	1.908	48.914	1.884	47.631	1.820	50.591	2.097	48.963	2.436	49.803	1.876	47.937	2.109
gZ	4.057	1.153	5.822	1.879	3.120	1.570	4.701	1.801	3.270	1.912	4.681	1.567	4.573	2.074	5.624	1.817
nX	0.210	0.427	0.173	0.437	-0.175	0.501	-0.035	0.466	-0.019	0.482	-0.117	0.408	-0.233	0.476	-0.197	0.410
nY	36.922	2.047	36.783	2.345	36.290	2.072	37.045	2.170	37.304	2.054	37.856	2.001	33.844	2.583	34.239	2.215
nZ	3.200	1.533	3.991	1.566	2.042	1.417	3.618	1.617	2.204	1.197	3.542	1.545	2.576	1.705	3.731	1.357
enLX	15.654	1.274	15.319	1.636	17.689	1.582	17.044	1.232	17.949	1.323	17.335	1.441	18.332	1.394	18.543	1.181
enLY	26.262	0.883	27.251	0.958	26.487	0.962	27.471	0.962	26.910	0.992	27.969	1.080	26.769	0.649	27.840	0.843
enLZ	-11.025	0.887	-12.397	0.912	-12.093	1.288	-13.102	1.178	-12.103	1.000	-13.150	0.980	-12.250	0.925	-13.037	1.093
enRX	-15.864	1.180	-15.900	1.334	-17.032	1.385	-16.525	1.218	-17.482	1.104	-17.550	1.395	-18.284	1.587	-18.428	1.750
enRY	26.357	1.203	27.338	1.071	26.304	1.001	26.881	0.997	26.661	0.937	28.119	0.831	27.096	0.796	27.873	0.752
enRZ	-11.051	0.872	-12.445	0.940	-11.305	1.173	-12.217	1.121	-11.479	0.937	-12.757	0.998	-12.044	1.154	-12.824	1.141
exLX	45.607	1.829	44.521	2.035	45.328	1.711	43.669	1.792	44.943	1.811	43.500	1.920	44.644	1.916	44.429	1.869
exLY	27.164	1.404	27.737	1.395	26.621	1.279	27.063	1.430	27.279	1.194	27.400	1.344	27.763	1.018	27.648	1.237
exLZ	-18.583	1.455	-19.149	1.528	-16.853	1.545	-17.837	1.181	-17.114	1.411	-17.327	1.405	-17.507	1.505	-18.109	1.164
exRX	-46.026	1.613	-44.861	2.329	-45.148	1.542	-43.239	1.941	-44.414	1.543	-42.974	1.676	-44.545	1.898	-44.107	1.864
exRY	27.583	1.200	28.245	1.761	26.529	1.295	27.742	1.329	27.147	1.267	27.798	1.323	28.662	1.146	28.394	1.352
exRZ	-19.410	1.343	-19.553	1.641	-16.927	1.327	-17.754	1.400	-16.722	1.366	-17.485	1.315	-17.452	1.776	-17.922	1.275

psLX	31.712	1.471	30.467	1.765	32.041	1.490	30.572	1.246	31.232	1.631	30.223	1.557	31.301	1.401	31.410	1.664
psLY	34.141	1.373	33.517	1.615	34.811	1.221	34.527	1.340	34.789	1.003	34.055	1.214	34.371	1.390	33.226	1.602
psLZ	-8.060	1.277	-8.083	1.226	-8.036	1.260	-8.007	1.541	-6.920	1.224	-7.336	1.256	-7.504	1.344	-7.051	1.132
psRX	-31.442	1.274	-30.110	1.750	-31.677	1.461	-30.194	1.302	-31.081	1.630	-30.149	1.470	-31.055	1.308	-30.991	1.691
psRY	34.433	1.360	33.902	1.559	35.003	1.058	34.497	1.200	34.701	1.278	34.497	1.221	34.853	1.082	33.743	1.425
psRZ	-7.830	1.164	-8.031	1.201	-7.790	1.443	-7.805	1.402	-6.690	1.052	-7.330	0.993	-7.227	1.270	-6.958	1.195
piLX	31.805	1.244	30.866	1.721	32.513	1.432	31.021	1.334	32.048	1.640	30.847	1.542	31.859	1.410	31.877	1.486
piLY	22.545	1.273	23.408	1.082	22.508	1.240	23.803	1.096	22.678	1.185	23.535	1.089	23.616	0.948	24.086	1.058
piLZ	-10.463	1.115	-11.751	0.999	-9.264	1.188	-10.950	1.226	-10.498	1.099	-11.961	1.114	-11.136	1.200	-12.480	0.754
piRX	-31.191	1.476	-30.045	1.961	-31.793	1.409	-30.447	1.354	-31.158	1.532	-30.352	1.540	-31.368	1.450	-31.200	1.766
piRY	22.905	1.424	23.448	1.159	22.930	1.089	23.938	0.978	22.742	1.204	23.975	1.057	23.923	1.041	24.235	1.211
piRZ	-10.348	1.018	-11.665	1.051	-9.041	1.040	-10.855	1.322	-10.346	1.162	-11.805	1.154	-11.212	1.213	-12.311	0.759
prnX	0.010	1.122	0.142	1.224	-0.098	0.867	-0.225	0.948	-0.225	0.780	0.299	1.104	-0.086	0.661	-0.161	0.933
prnY	-3.488	1.600	-4.562	2.254	-2.347	1.739	-2.576	1.826	-3.475	2.418	-3.443	1.743	-3.030	2.055	-2.670	1.697
prnZ	27.416	1.870	28.796	25.773	1.798	1.786	26.841	1.653	25.834	1.578	27.174	1.466	25.092	1.875	26.073	2.121
snX	0.081	0.661	0.090	0.601	-0.090	0.606	-0.212	0.624	-0.116	0.551	0.152	0.704	0.003	0.510	-0.201	0.543
snY	-16.517	1.458	-16.496	2.067	-16.469	2.189	-15.810	1.522	-16.969	1.897	-16.248	2.013	-15.956	1.614	-15.303	1.349
snZ	11.450	1.491	12.022	1.584	10.576	1.160	11.576	1.535	10.917	1.101	11.891	1.494	10.572	1.264	11.225	1.607
alLX	16.114	1.058	17.303	1.346	16.828	1.111	16.709	1.302	16.833	1.462	16.930	0.920	16.665	1.171	16.616	1.202
alLY	-7.232	1.163	-6.755	1.500	-6.175	1.381	-6.286	1.412	-6.573	1.448	-6.642	1.782	-6.549	1.144	-6.010	1.486
alLZ	6.724	1.898	7.162	1.793	5.045	1.346	5.423	1.457	4.632	1.441	5.281	1.183	5.682	1.352	5.879	1.632
alRX	-16.558	1.155	-17.863	1.284	-17.464	1.233	-17.454	1.289	-17.298	1.372	-17.419	1.346	-16.786	1.079	-17.353	1.072
alRY	-7.269	1.158	-6.951	1.313	-6.015	1.278	-6.366	1.180	-6.618	1.317	-6.947	1.394	-6.520	1.454	-6.214	1.188
alRZ	7.234	1.939	7.515	1.895	5.353	1.264	5.710	1.642	4.755	1.363	5.492	1.342	5.968	1.347	6.182	1.944
lsX	0.170	0.511	0.088	0.541	-0.205	0.539	-0.205	0.446	-0.337	0.542	-0.186	0.555	-0.100	0.452	-0.023	0.475
lsY	-30.686	1.143	-31.144	1.112	-30.065	1.372	-30.582	1.311	-31.174	0.842	-31.549	1.116	-30.957	1.105	-31.629	1.419
lsZ	11.744	1.022	11.550	1.053	12.223	1.155	12.381	1.158	12.283	0.893	12.185	0.944	12.875	0.956	12.977	1.341
liX	0.145	0.611	0.397	0.460	0.068	0.414	0.137	0.363	0.304	0.599	0.029	0.419	0.128	0.549	0.212	0.505

liY	-45.936	1.434	-45.443	1.900	-46.105	1.503	-46.746	2.063	-45.832	1.837	-46.428	2.003	-45.382	2.032	-45.595	1.613
liZ	8.707	1.468	8.717	1.380	8.789	1.111	8.901	1.616	9.294	1.160	8.366	1.781	9.701	1.069	9.183	1.164
cphLX	6.173	0.836	6.208	1.164	6.108	1.194	6.431	1.092	5.868	0.881	5.789	0.977	5.609	0.806	5.998	0.846
cphLY	-27.559	1.142	-28.048	1.110	-28.186	1.383	-28.606	1.239	-28.447	0.966	-29.086	0.942	-28.784	1.126	-28.698	1.527
cphLZ	10.810	0.953	10.777	0.794	11.212	1.147	11.428	1.056	11.153	0.817	11.410	0.889	11.873	0.791	11.937	1.095
cphRX	-6.174	0.839	-6.648	1.285	-6.394	1.017	-6.867	0.941	-6.843	1.072	-6.440	0.980	-6.249	0.815	-6.478	0.654
cphRY	-27.595	1.109	-28.094	1.219	-27.997	1.296	-28.574	1.095	-28.275	0.869	-28.896	0.908	-28.672	1.094	-28.705	1.545
cphRZ	11.041	0.922	10.897	0.918	11.293	1.115	11.485	1.000	11.139	0.758	11.331	0.871	11.724	0.929	11.963	1.063
chLX	23.968	2.071	23.430	1.783	27.367	1.878	26.312	1.952	25.333	1.837	24.340	1.856	24.916	2.342	24.597	1.712
chLY	-38.020	1.197	-37.763	1.301	-38.300	1.185	-38.378	1.431	-38.673	1.098	-39.387	1.578	-39.600	1.197	-38.734	1.249
chLZ	-5.565	2.208	-4.580	1.833	-4.056	1.231	-3.460	1.754	-4.910	1.500	-3.367	1.267	-4.223	2.176	-3.818	1.596
chRX	-24.892	1.824	-24.179	1.656	-27.718	2.270	-26.744	2.156	-25.085	2.016	-23.956	1.616	-24.331	2.272	-24.431	2.311
chRY	-38.075	1.196	-37.472	1.354	-37.792	1.262	-38.322	1.319	-38.466	1.440	-38.736	1.457	-39.132	1.391	-39.004	1.136
chRZ	-5.413	2.081	-4.810	1.897	-3.894	1.354	-3.320	1.600	-4.510	1.618	-3.058	1.448	-3.950	1.685	-3.603	2.115
pgX	0.328	0.775	0.361	0.573	0.289	0.818	0.529	0.695	-0.160	0.917	0.070	0.761	0.058	0.906	0.135	0.588
pgY	-66.553	1.824	-68.002	2.327	-66.948	2.268	-68.351	2.612	-66.298	2.758	-66.803	2.857	-66.117	2.016	-66.658	2.057
pgZ	5.366	2.214	5.214	2.201	3.831	2.257	3.243	3.204	5.811	2.879	4.223	2.348	3.869	2.708	3.339	2.411

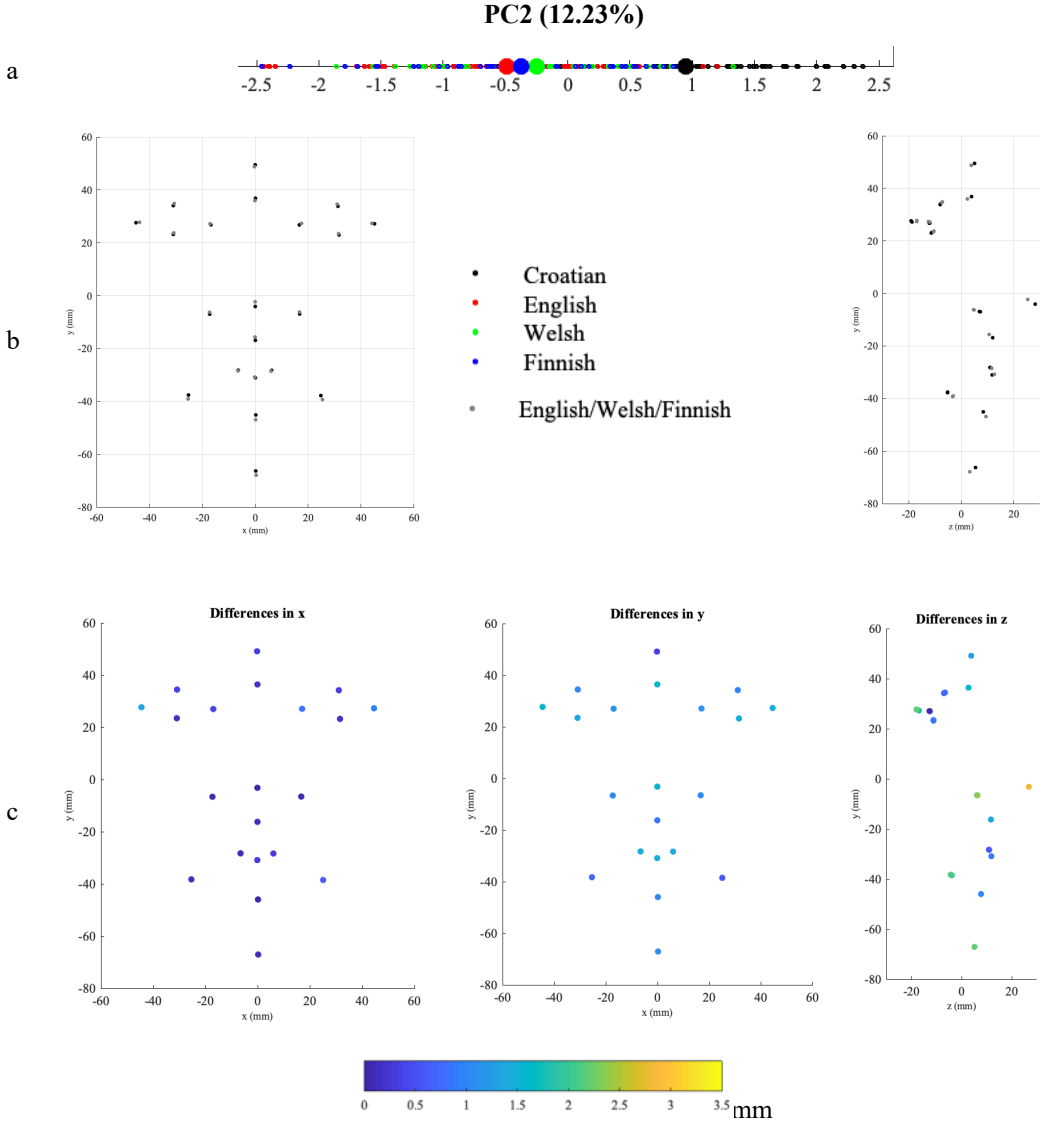
13 APPENDIX 3: SUPPLEMENTARY PLOTS

13.1 CONVENTIONAL PCA: GEOGRAPHICAL LOCATION



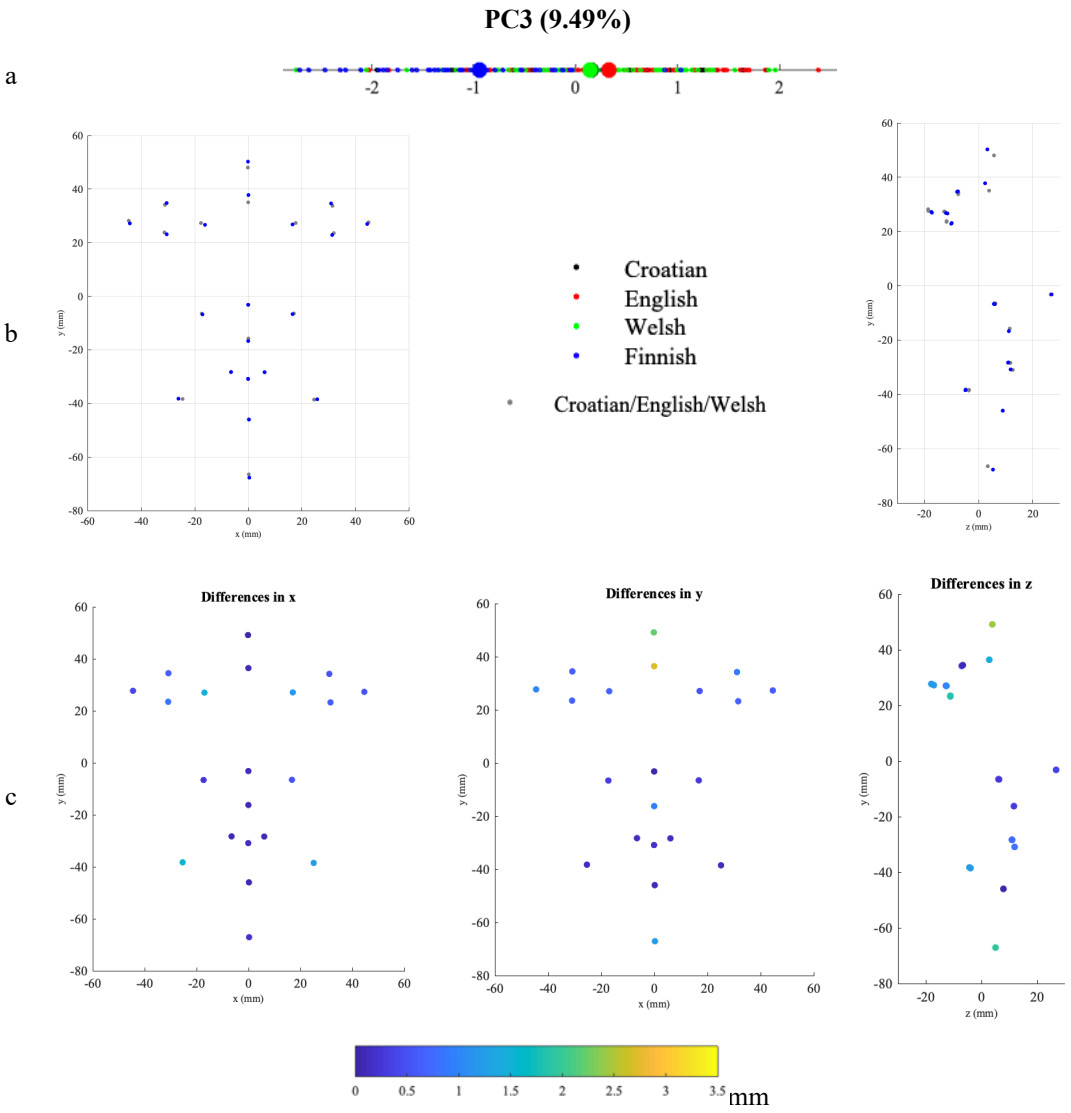
Main differences (English v Croatian/Welsh)

- English: wider and more prominent eyes (3mm), wider mouths (2.5mm)
- Croatian/Welsh: more prominent pronasale (2mm)



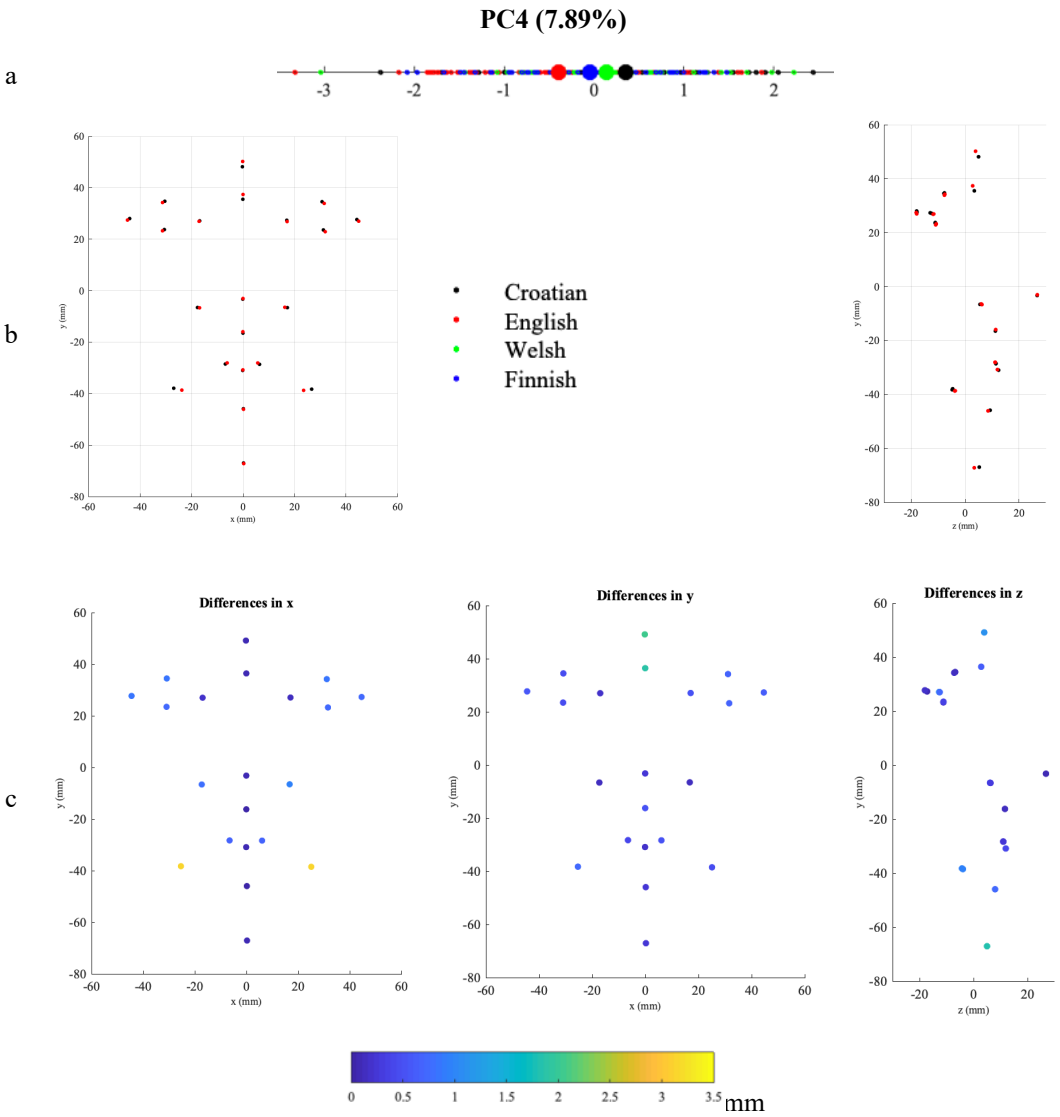
Main differences (English/Welsh/Finnish v Croatian)

- Croatian: more prominent pronasale (3mm), subnasale (2.5mm) and pogonion (2.5mm), more deep-set eyes (2.5mm)



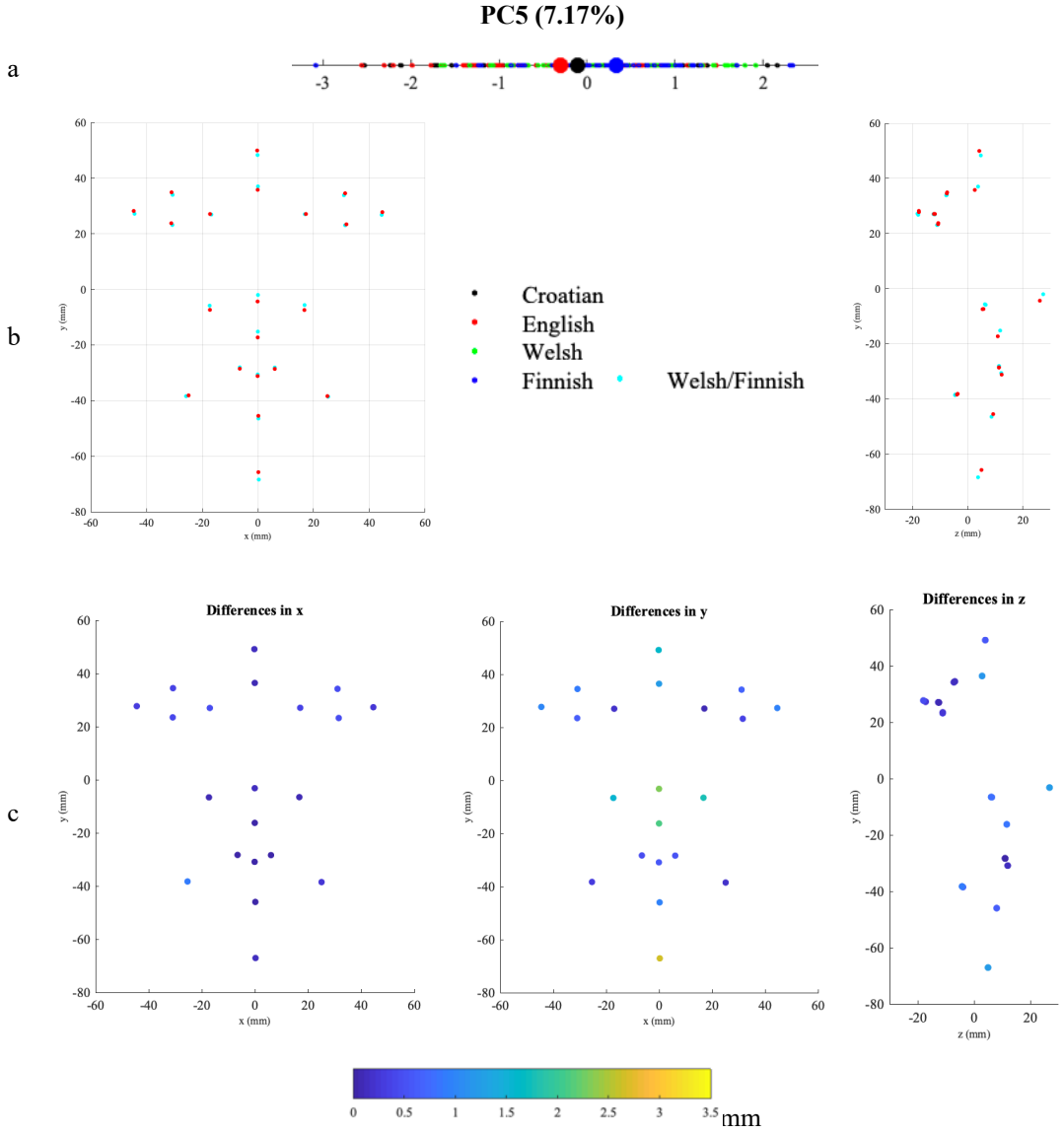
Main differences (Finnish v Croatian/English/Welsh)

- Finnish: more superior nasion (3mm) and glabella (2mm), less prominent glabella (2.5mm)



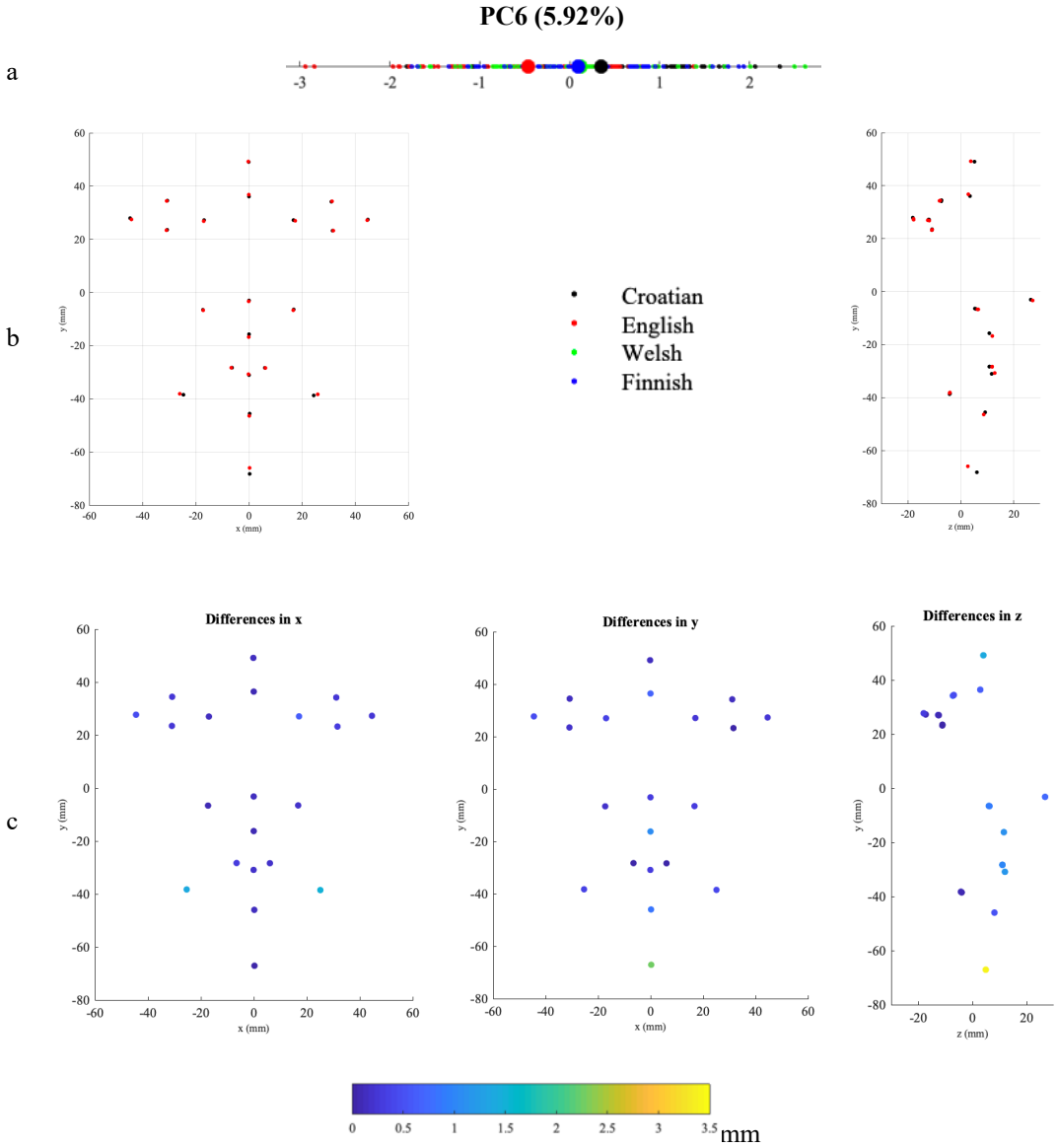
Main differences (English v Croatian)

- Croatian: wider mouth (3.5mm), more inferior nasion and glabella (2mm)
- English: narrower mouth (3.5mm), more superior nasion and glabella (2mm)



Main differences (English v Welsh/Finnish)

- English: more superior pogonion (2mm), more inferior pronasale (1.75mm)

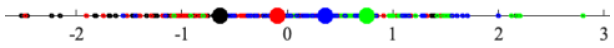


Main differences (English v Croatian)

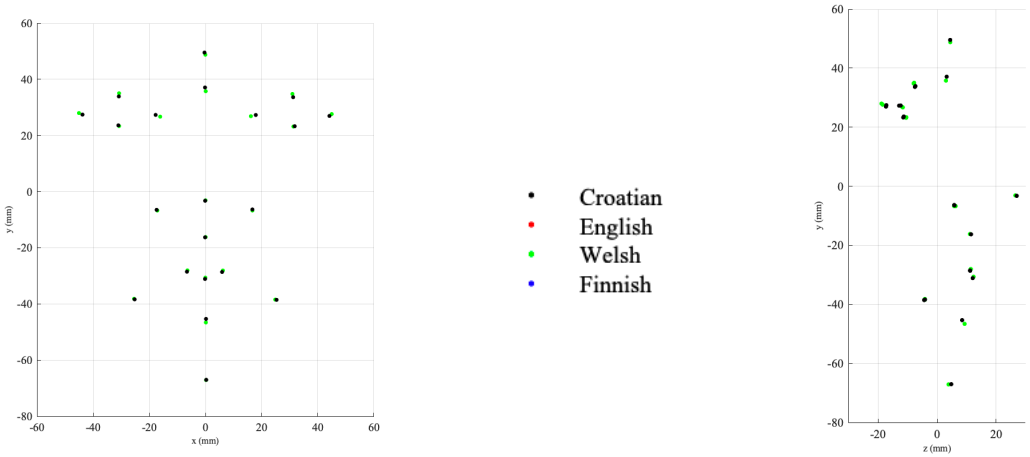
- Croatian: more prominent pogonion (3.5mm) and more inferior pogonion (2mm)
- English: less prominent pogonion (3.5mm) and more superior pogonion (2mm)

PC7 (4.54%)

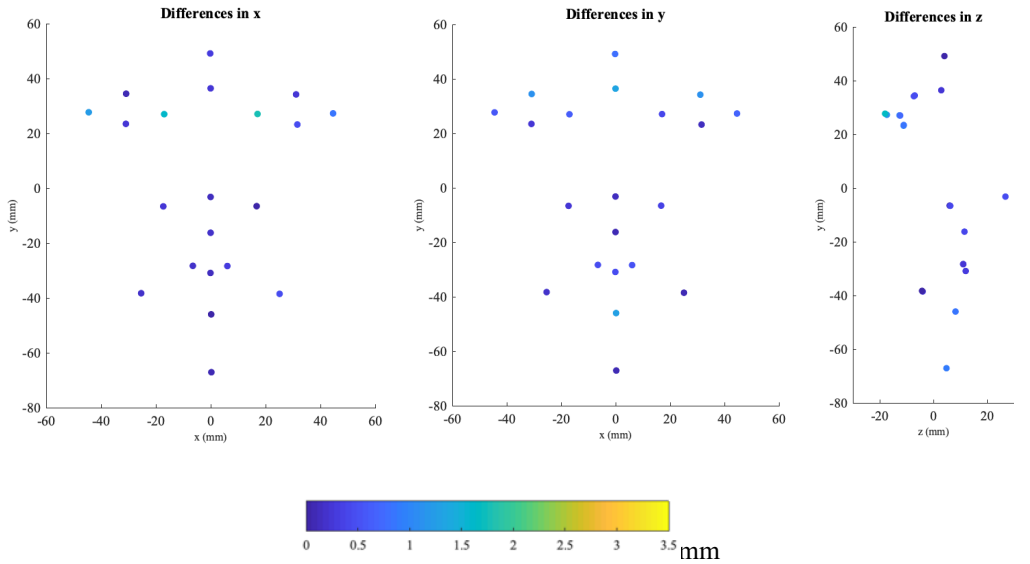
a



b



c

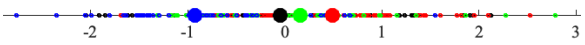


Main differences (Croatian v Welsh)

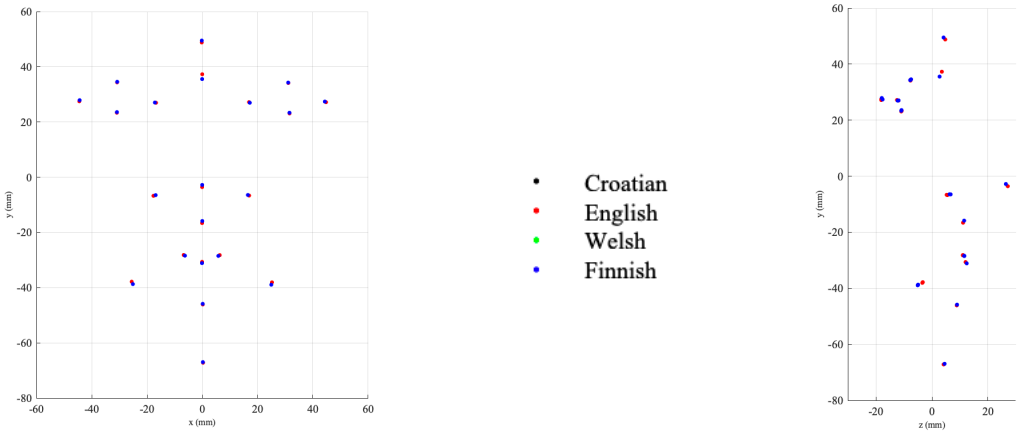
- Croatian: narrower eyes (1.5mm)
- Welsh: wider eyes (1.5mm)

PC8 (3.26%)

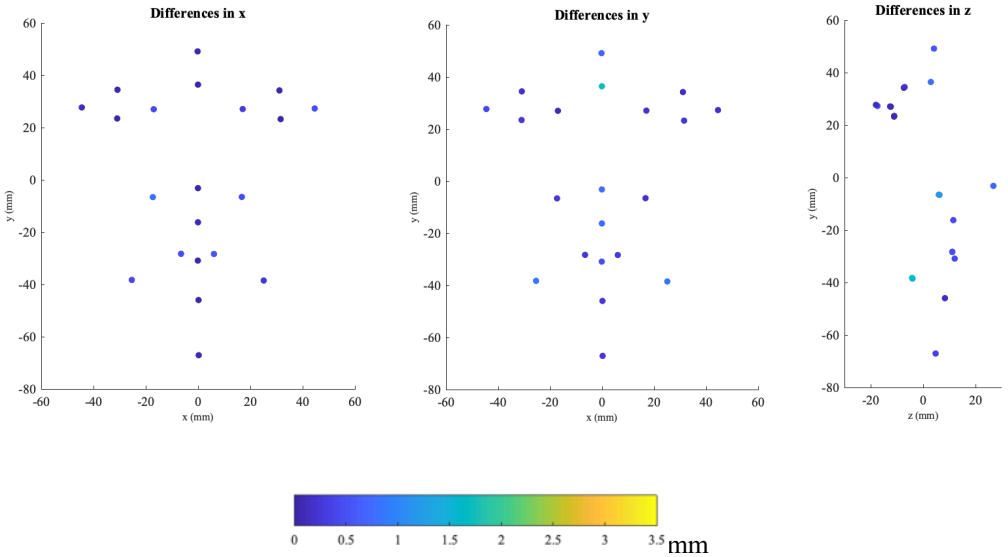
a



b

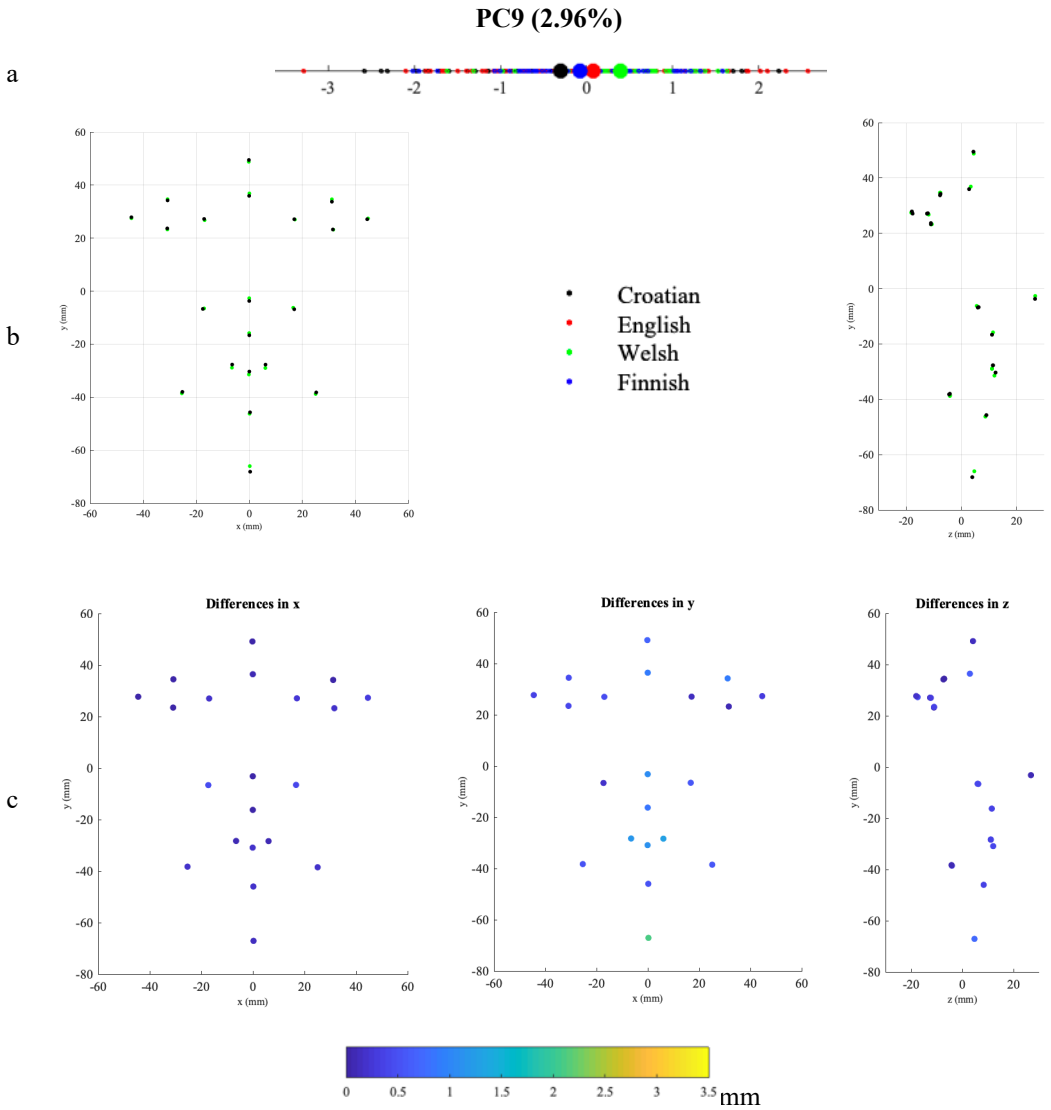


c



Main differences (Finnish v English)

- English: more prominent corners of mouth (1.5mm), more superior nasion (1.5mm)
- Finnish: more inset corners of mouth (1.5mm), more inferior nasion (1.5mm)



Main differences (Croatian v Welsh)

- Croatian: more inferior pogonion (2mm)
- Welsh: more superior pogonion (2mm)

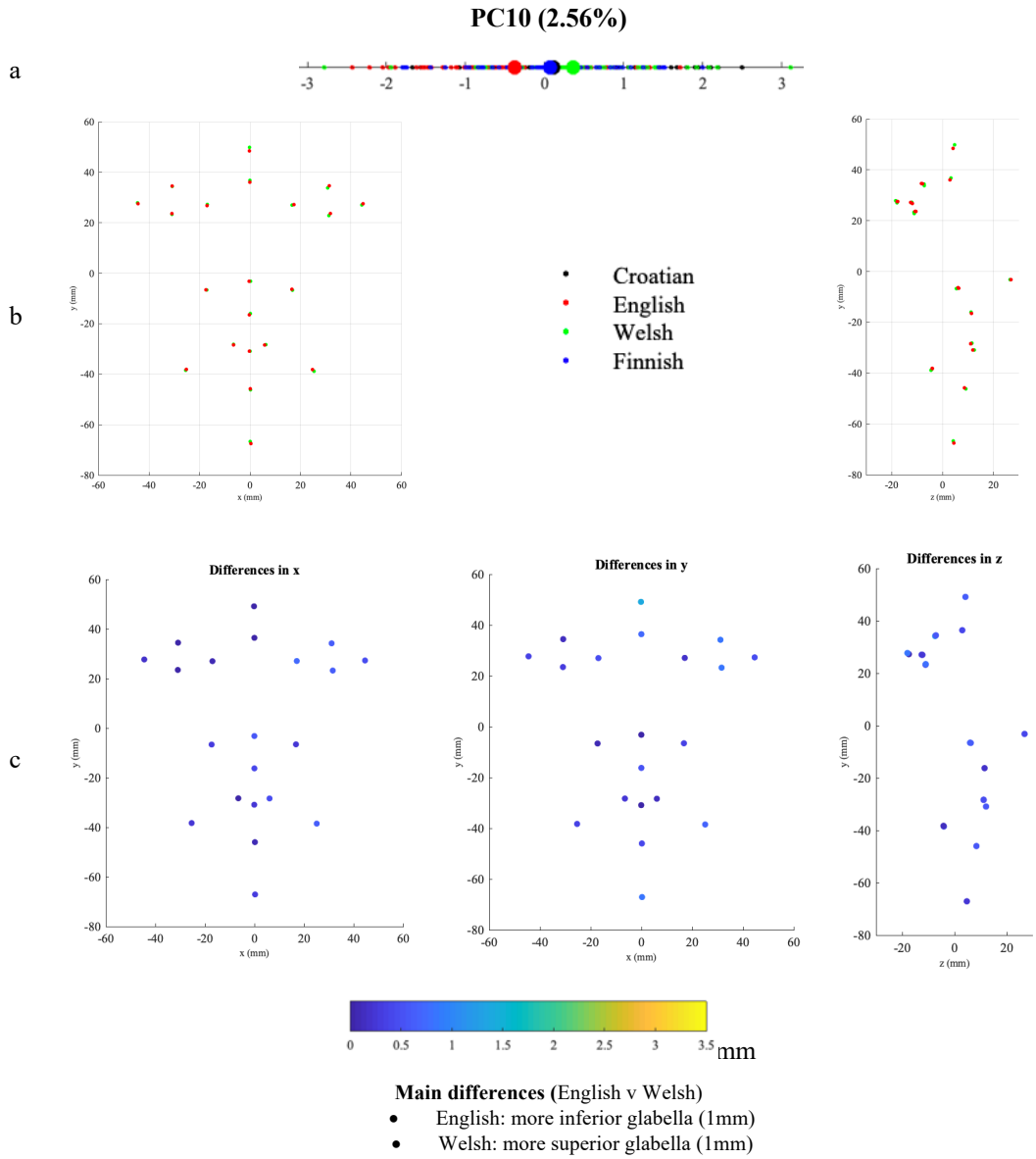
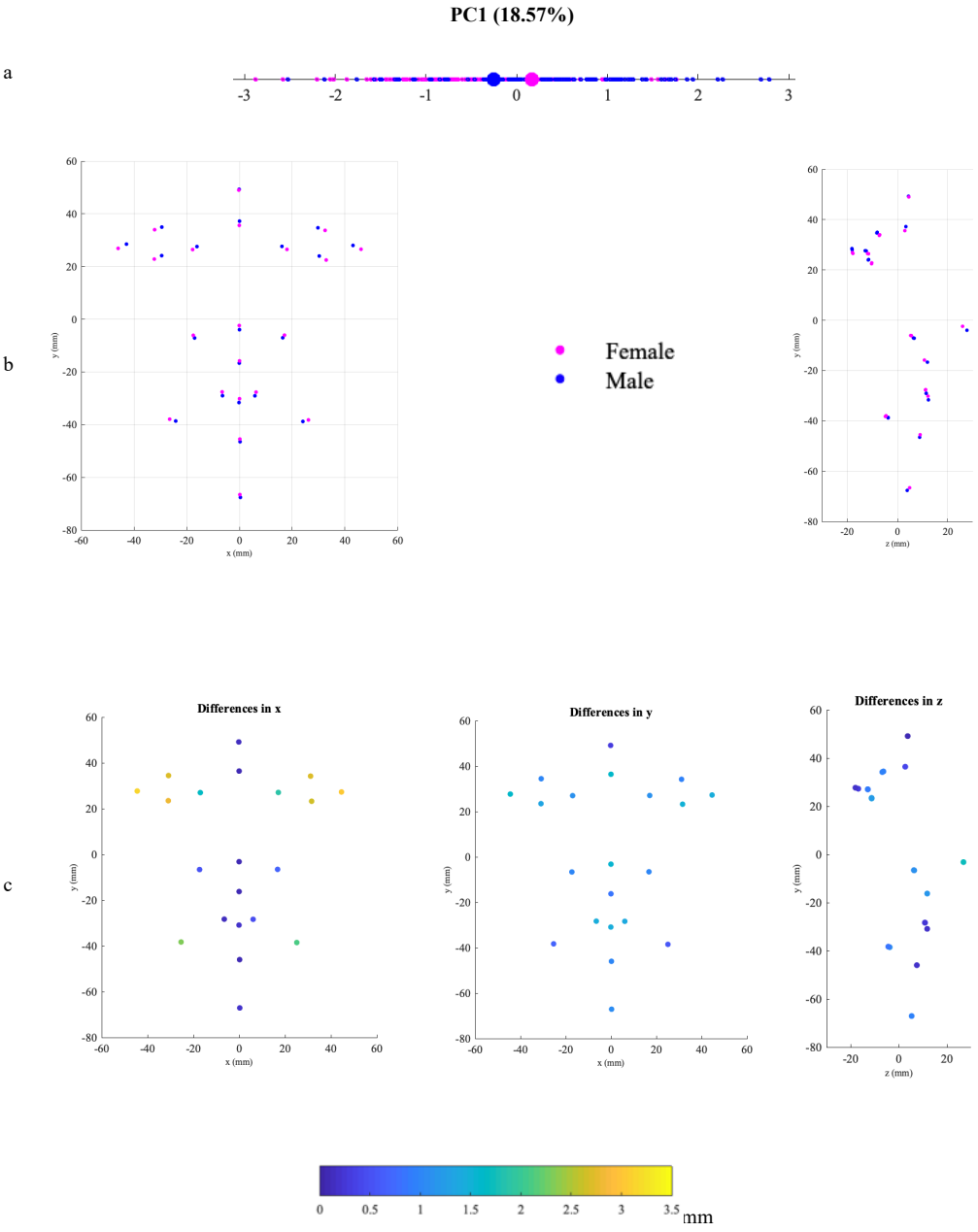


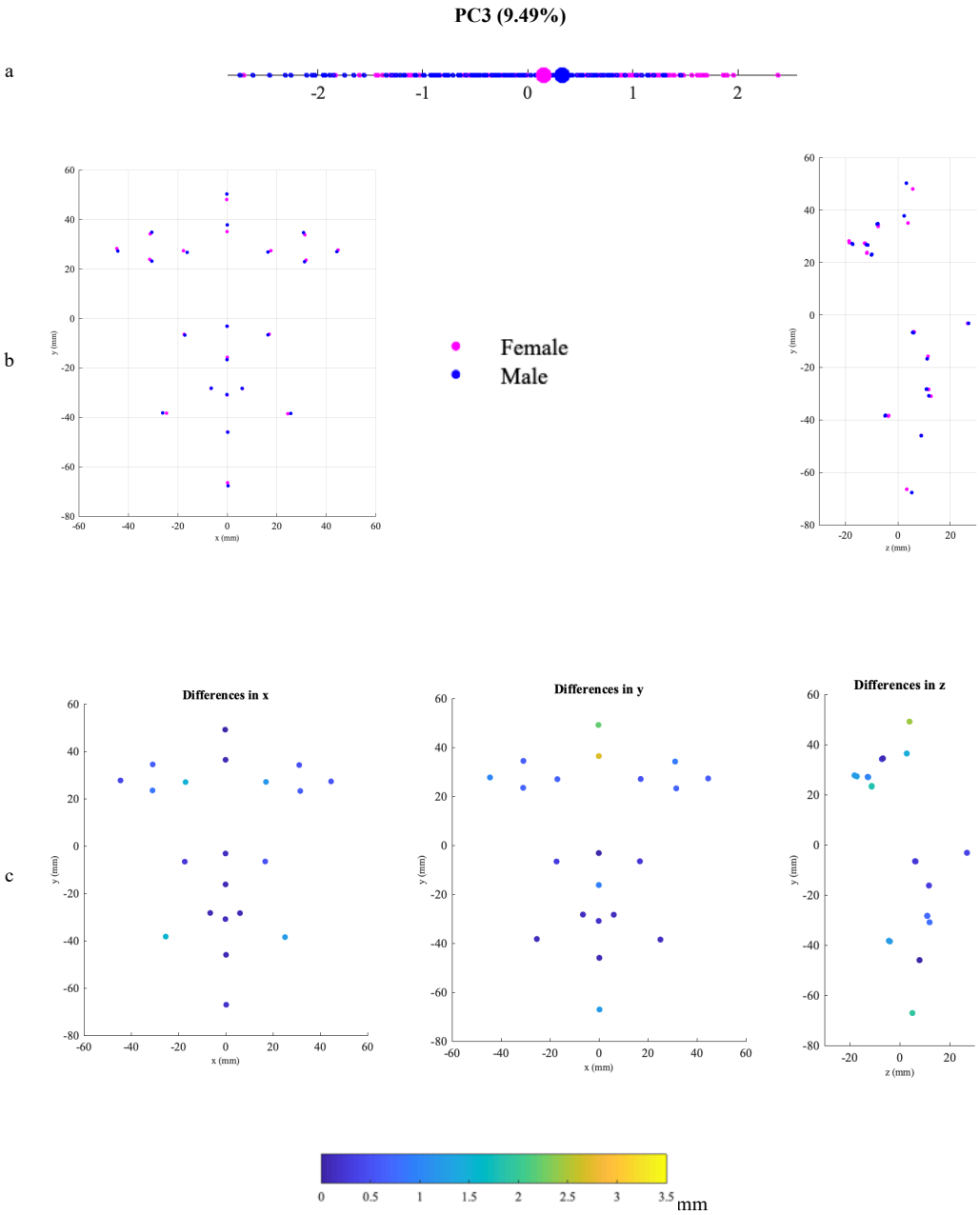
Figure 78: Conventional PC1-10 (grouped by geographical location). Significant following univariate ANOVA (p<0.003 to account for Bonferroni correction). PC1 did not reach significance but is included to demonstrate a contradiction in the findings between the PC1 and PC4. (a) Standardised component scores, (b) Mean face +/- square root of the respective eigenvalue/eigenvector, (c) Difference between mean + sqrt(eigenvalue)*eigenvector and mean - sqrt(eigenvalue)*eigenvector (mm).

13.2 CONVENTIONAL PCA: SEX



Main differences

- Females: wider eyes (3mm), wider mouths (2.5mm), less prominent pronasale (2mm)



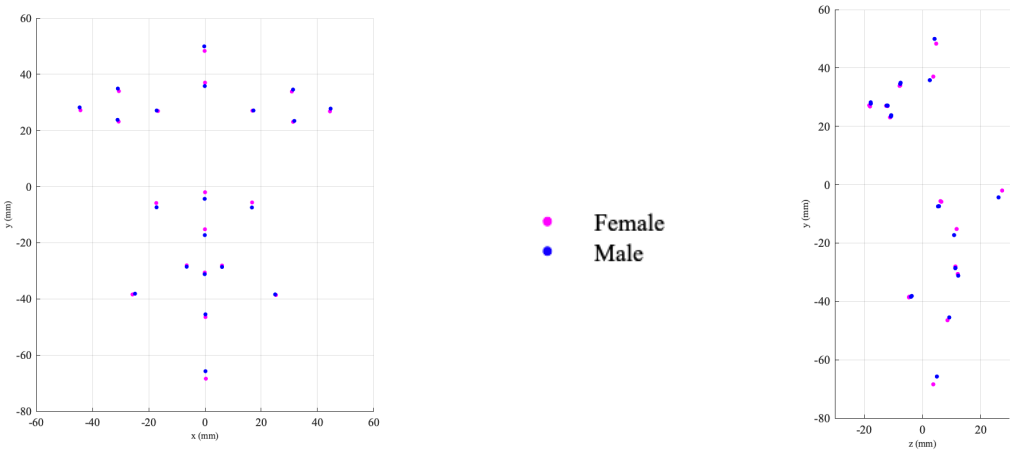
Main differences

- Females: more inferior nasion (3mm), more inferior (2mm) and less prominent glabella (2.5mm), less prominent pogonion (2mm)

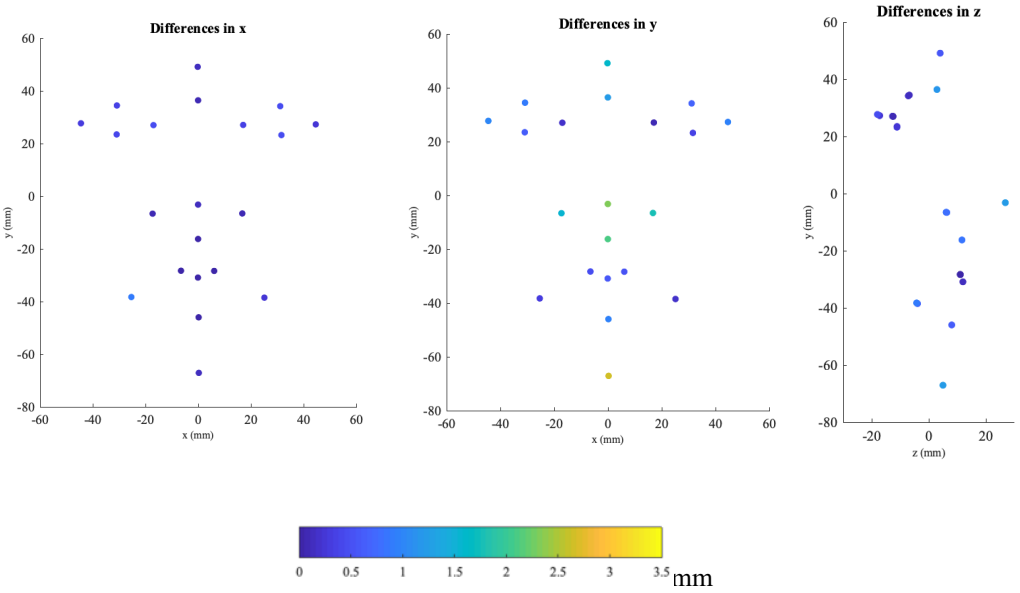
PC5 (7.17%)

a

b



c

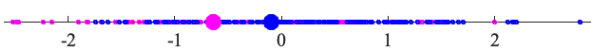


Main differences

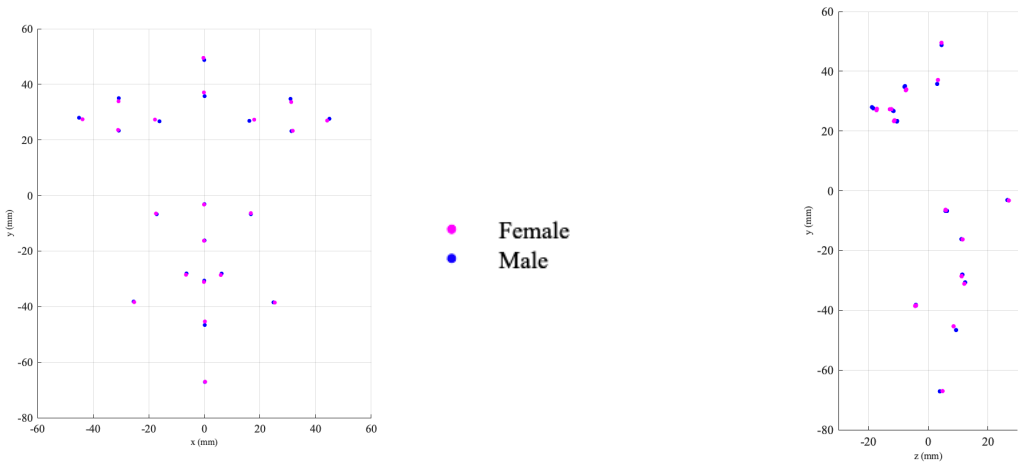
- Females: more inferior pogonion (3mm), more superior pronasale (2mm)

PC7 (4.54%)

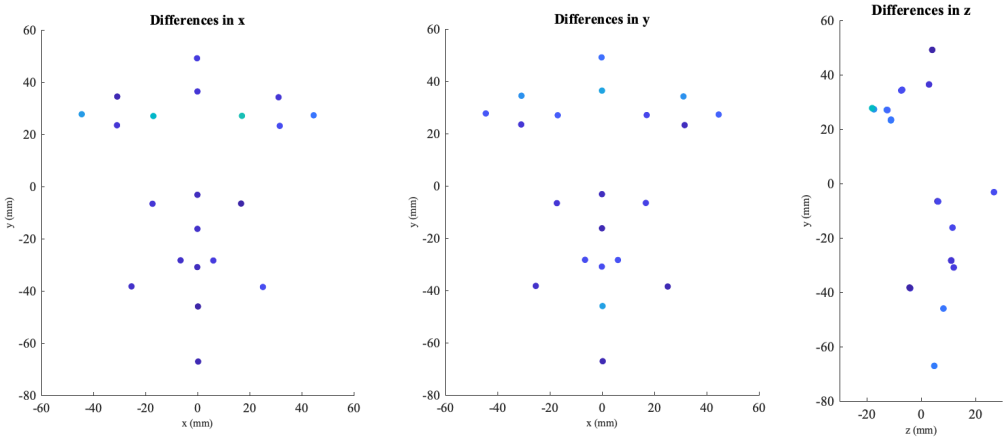
a

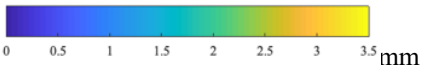


b



c





Main differences

- Females: eyes wider apart (1.5mm)

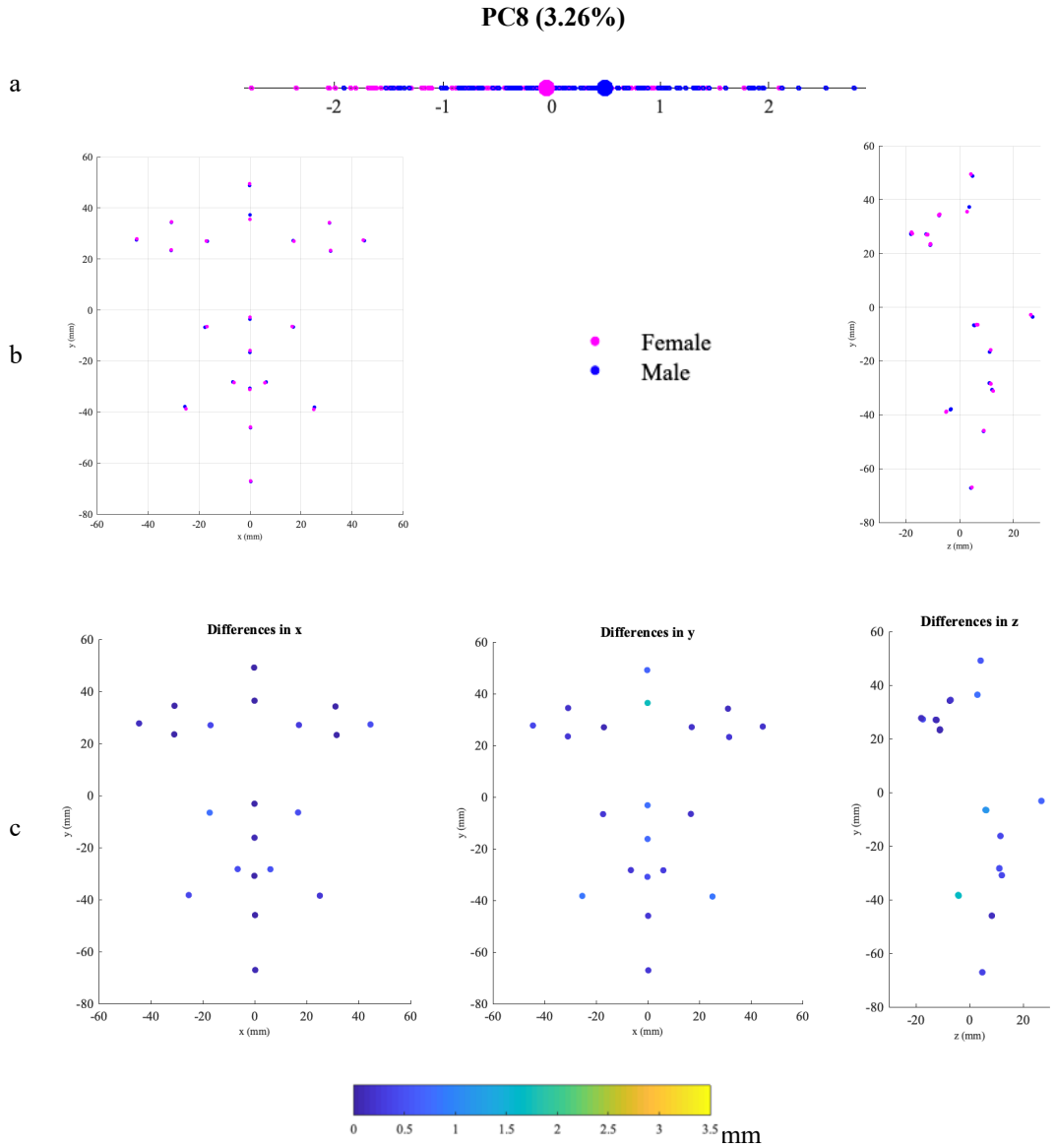


Figure 79: Interpretation of conventional PC1, 3, 5, 6 and 8 (grouped by sex). Significant following univariate ANOVA (p<0.003 to account for Bonferroni correction). (a) Standardised component scores, (b) Mean face +/- respective eigenvalue/eigenvector, (c) Difference between mean + eigenvalue/eigenvector and mean - eigenvalue/eigenvector (mm).

13.3 MULTILEVEL PRINCIPAL COMPONENT ANALYSIS: GEOGRAPHICAL LOCATION

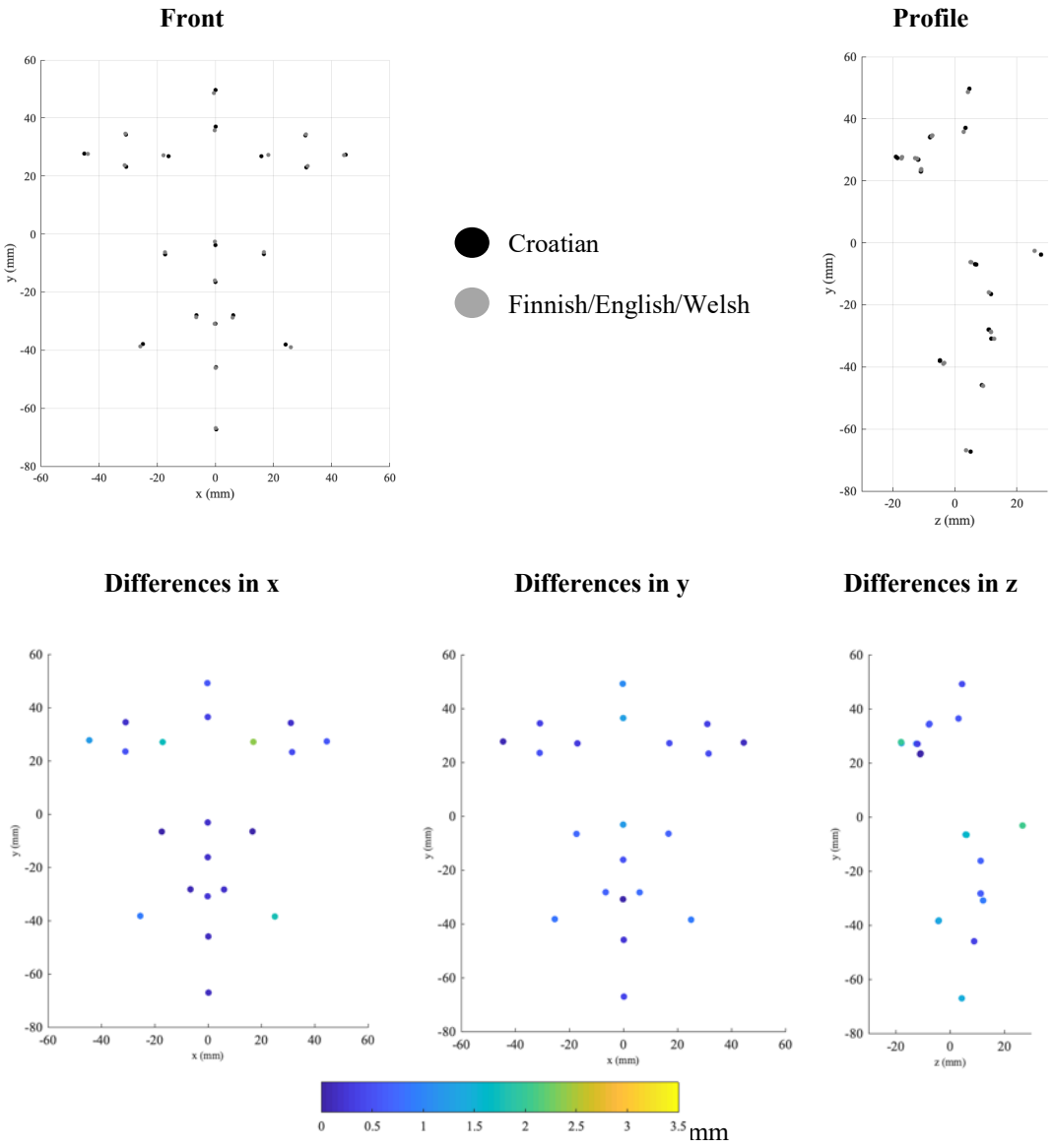


Figure 80: The facial differences along mPCA geographical location PC1 – Croatians included in model

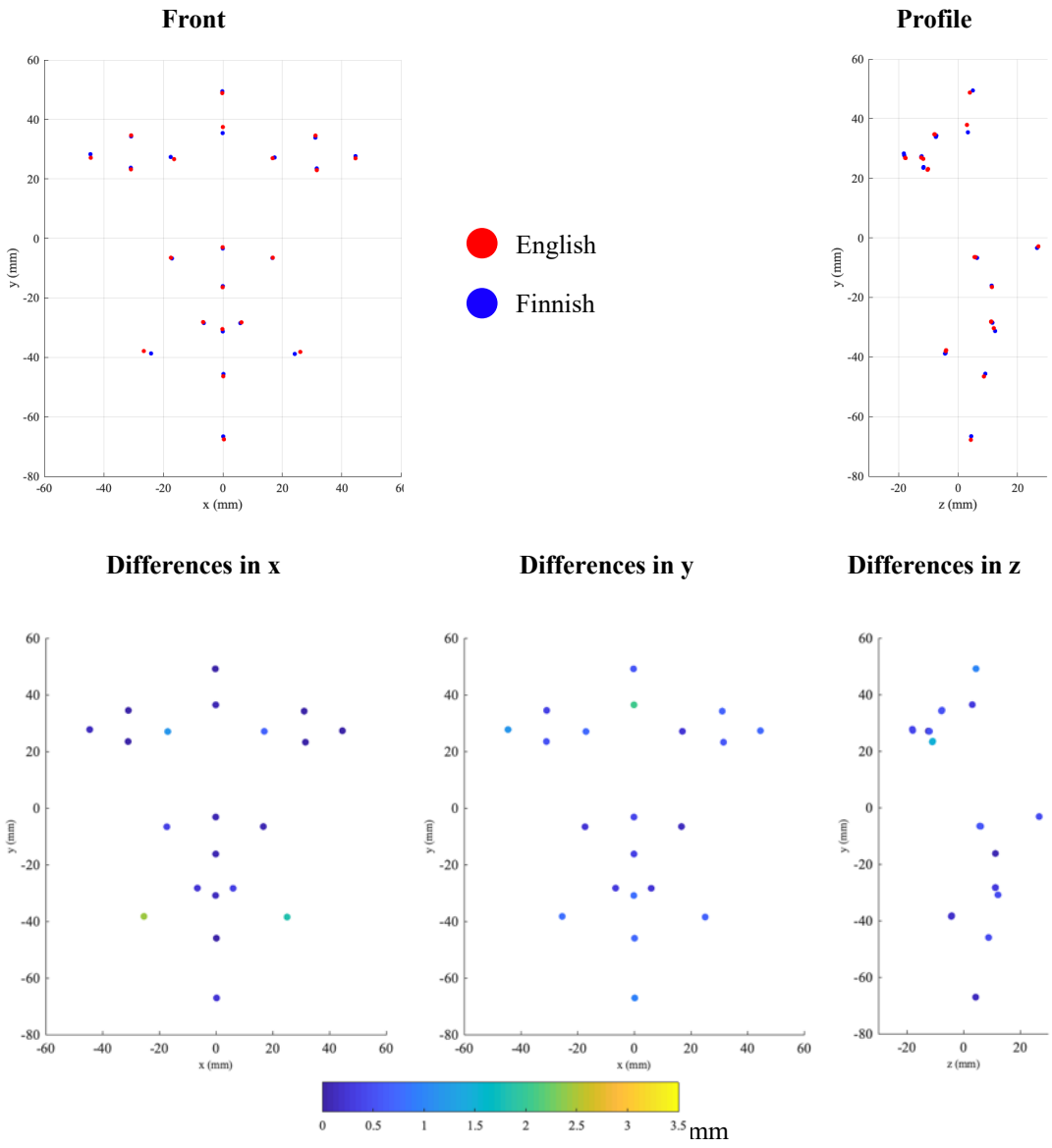


Figure 81: The facial differences along mPCA geographical location PC2 – Croatians included in model

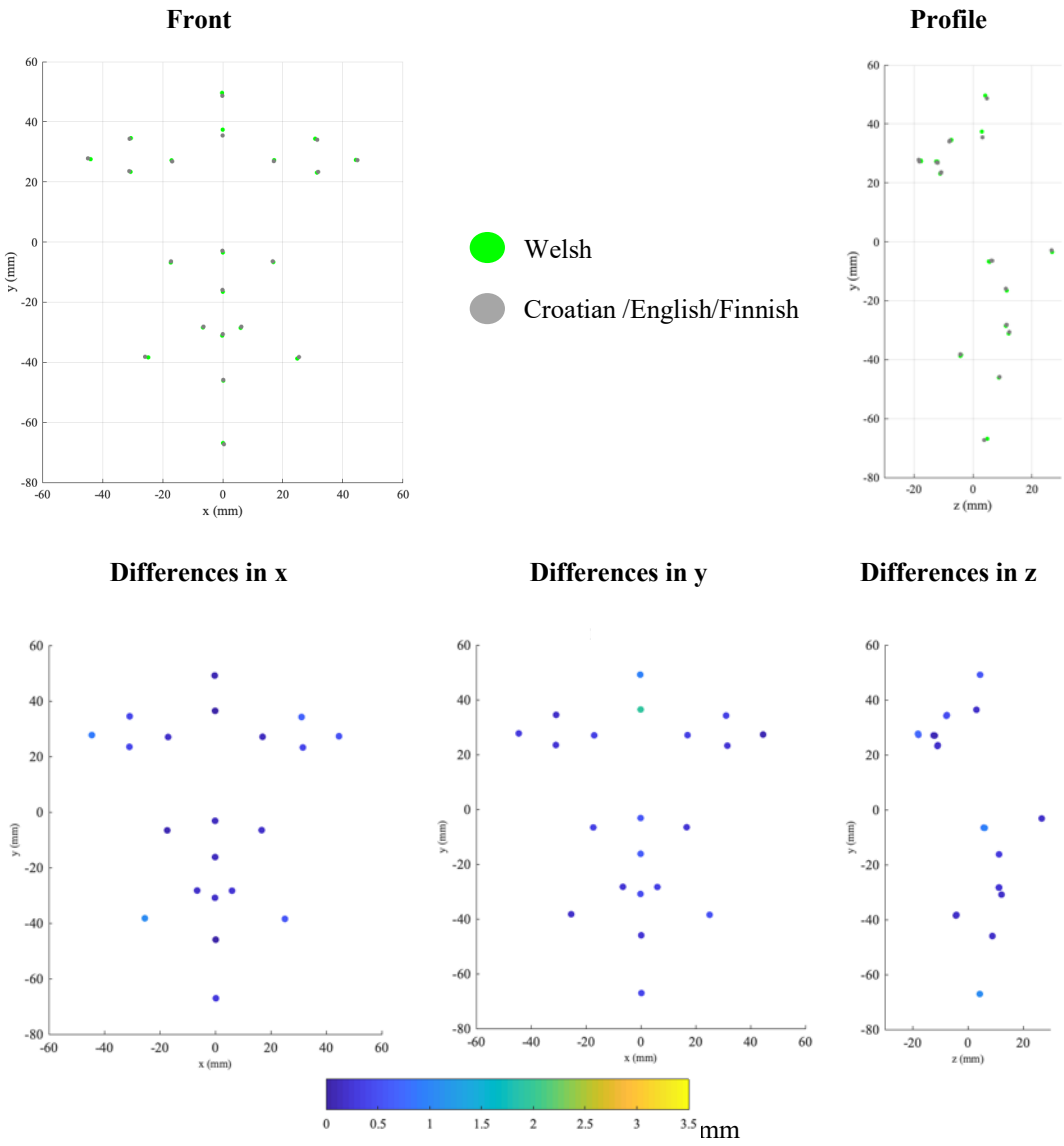


Figure 82: The facial differences along mPCA geographical location PC3 – Croatians included in model

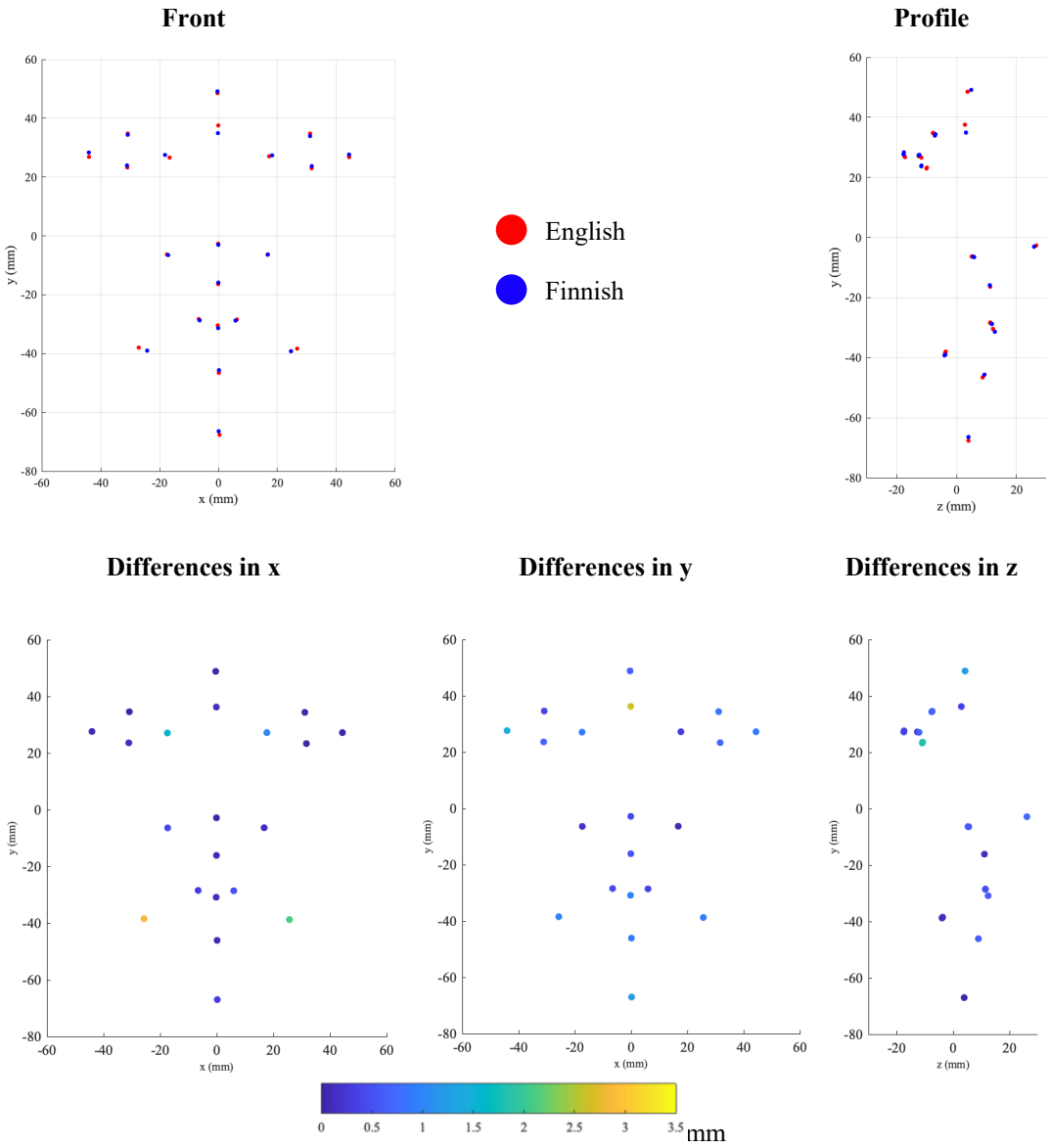


Figure 83: The facial differences along mPCA geographical PC1 – Croatians excluded from model

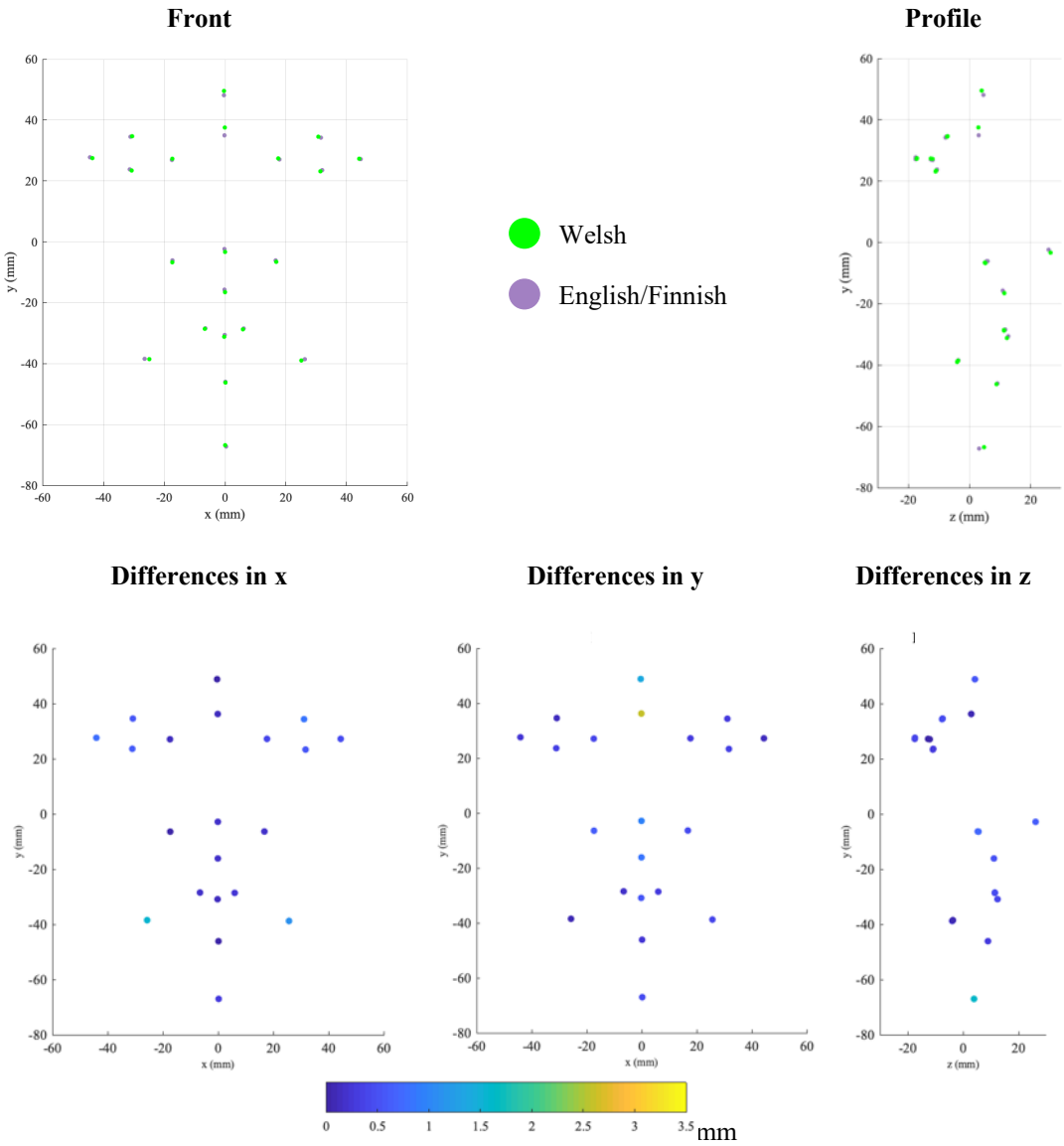


Figure 84: The facial differences along mPCA geographical location PC2 – Croatians excluded from model

13.4 MULTILEVEL PRINCIPAL COMPONENT ANALYSIS: SEX

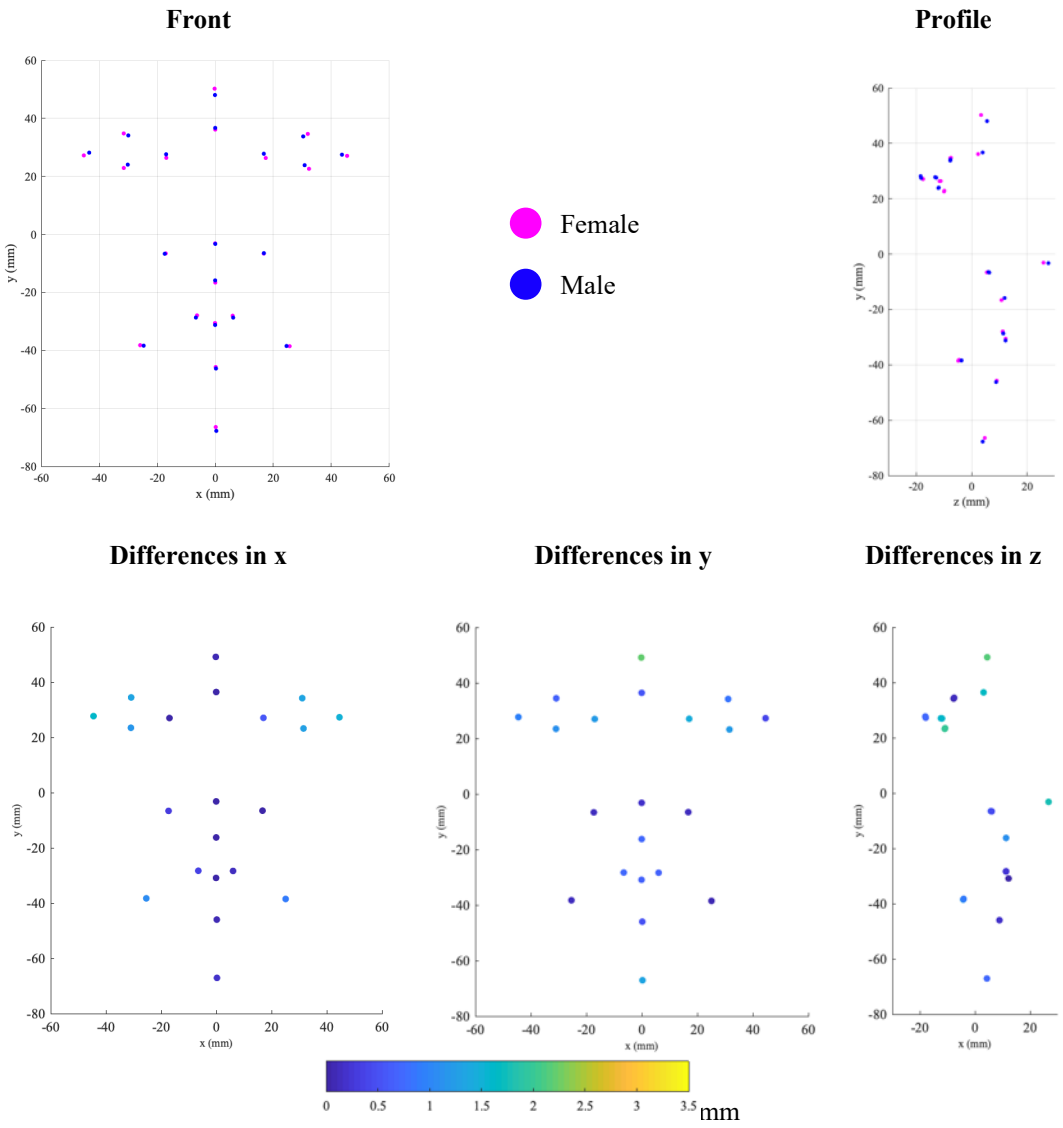


Figure 85: The facial differences along mPCA sex PC1 – Croatians included in the model

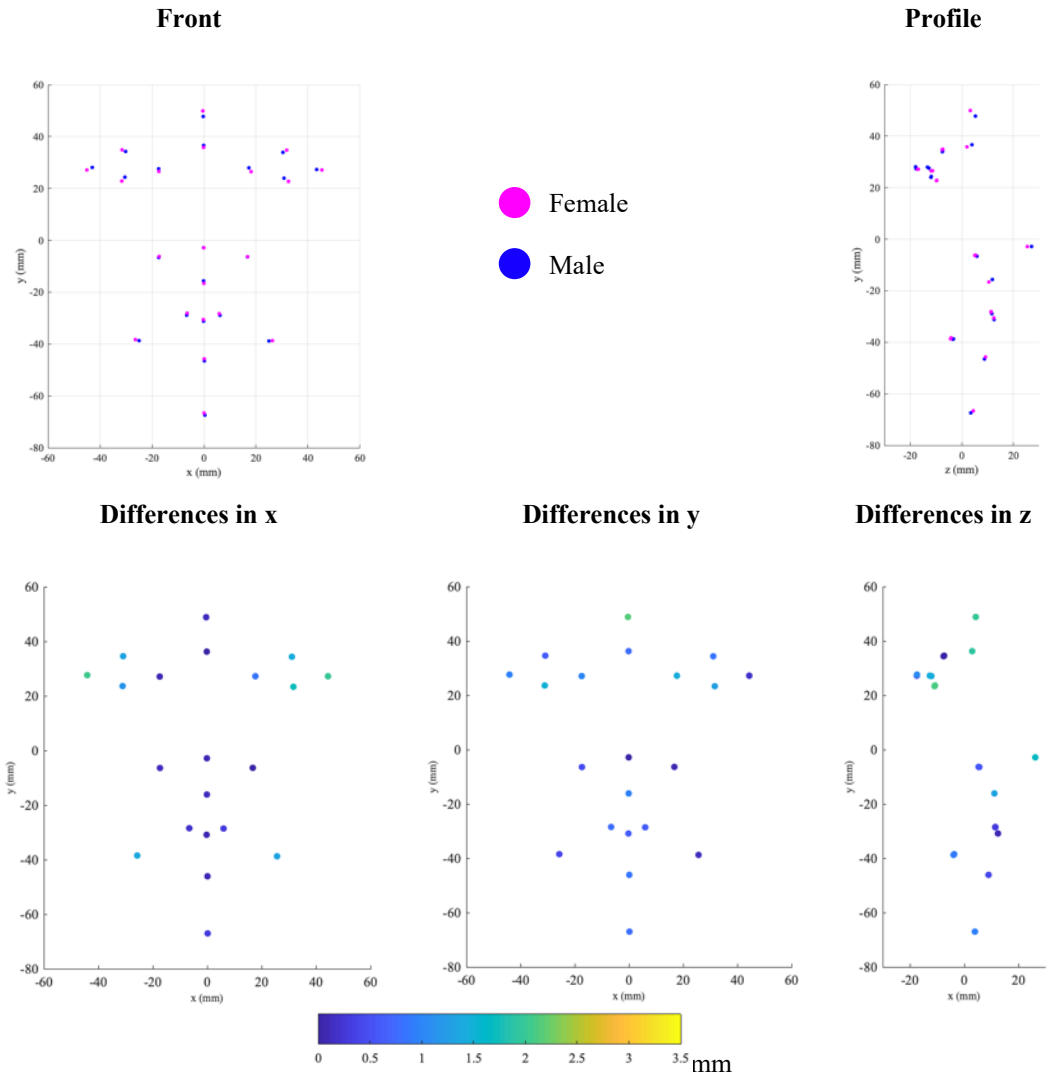


Figure 86: The facial differences along mPCA sex PC1 – Croatians excluded from the model