

# FACE SHAPE ANALYSIS IN PEOPLE WITH EPILEPSY

**Vamsi Krishna Chinthapalli**

**UCL Institute of Neurology**

**A thesis submitted to UCL for the degree of Doctor of Philosophy**

## DECLARATION

**I, Vamsi Krishna Chinthapalli, confirm that the work presented in this document is my own. Where information has been derived from other sources, I confirm that this has been indicated in the document.**

**Signature:**

**Date:                    19<sup>th</sup> February 2018**

## ABSTRACT

Stereophotogrammetry and dense surface modelling are novel techniques that have been used to study face shape in genetic and neurodevelopmental disorders. In people with epilepsy, it has been recognised that the condition may be associated with underlying structural variants or malformations of cortical development in some cases. Here I recruited 869 people with epilepsy or unaffected relatives and control subjects to study face shape. I sought to explore whether face shape and symmetry, using new metrics for each, could help to predict those people with epilepsy who may have potential underlying genetic or structural causes.

My reproducibility studies found that stereophotogrammetry and dense surface modelling were susceptible to error from changes in head position or face expression, but not from camera calibration, image acquisition and image landmarking. The next study found that in people with epilepsy, a measurement of atypical face shape, Face Shape Difference (FSD), was significantly increased in those with pathogenic structural variants compared to those without pathogenic structural variants. The FSD value was used to predict the presence of pathogenic structural variants with a sensitivity of 66-80% and specificity of 65-78%. Body mass index affects face shape in a partly predictable manner. The effect of body mass index differences was controlled for in a further analysis. I then analysed facial asymmetry and showed that it was increased in people with developmental lesions in the brain but not in people with pathogenic structural variants. A final study showed that stereophotogrammetry, dense surface modelling, FSD and reflected FSD could be used to study a single genetic disorder associated with epilepsy, to find previously unrecognised face shape changes. Stereophotogrammetry and dense surface modelling therefore appear to be promising tools to aid both in discovery of underlying causes for epilepsy and in understanding of such causes in terms of facial development.

# TABLE OF CONTENTS

Declaration.....	2
Abstract.....	3
List of Abbreviations.....	11
List of Figures.....	13
List of Tables.....	15
Acknowledgements.....	16
Author's contribution.....	17
Publication.....	18
Chapter 1: Introduction and literature review.....	19
1.1 The study of face shape.....	19
1.1.1 Aristotle and physiognomy.....	19
1.1.2 Gall and phrenology.....	21
1.1.3 Lombroso and epilepsy.....	23
1.1.4 Dysmorphology.....	24
1.2 The study of face shape measurement.....	28
1.2.1 Direct anthropometry.....	28
1.2.2 Indirect anthropometry (two dimensional).....	32
1.2.3 Indirect anthropometry (three-dimensional).....	34
1.2.4 Computed tomography and magnetic resonance imaging.....	36
1.2.5 Stereophotogrammetry.....	37
1.2.6 Methods of analysing 3D face shapes.....	40
1.2.7 Dense surface modelling using stereophotogrammetry.....	45
1.3 Face shape and epilepsy.....	47
1.3.1 Brain and face development.....	47
1.3.2 Epilepsy.....	50
1.3.3 Genetic or developmental disorders causing epilepsy and facial dysmorphism.....	51
1.3.4 Genetic epilepsy.....	55
1.3.5 Benefits of genetic diagnosis.....	59
1.3.6 Trauma.....	60
1.3.7 Anti-epileptic drugs.....	61
1.4 Aims.....	63
1.4.1 Validity of stereophotogrammetry and dense surface modelling.....	63

1.4.2 Investigation of face shape in people with epilepsy.....	63
1.5 Experimental approach .....	65
1.6 Justification of research.....	67
Chapter 2: Methods .....	69
2.1 Literature review .....	69
2.2 Recruitment of subjects.....	69
2.2.1 Reference subjects.....	69
2.2.2 Recruitment of control subjects .....	70
2.2.3 Recruitment of people with epilepsy in the United Kingdom .....	71
2.2.4 Recruitment of people with epilepsy in Europe .....	71
2.2.5 Groups used for analysis.....	72
2.3 Image capture .....	79
2.4 Other data collection.....	84
2.5 Landmarking.....	85
2.5.1 Exocanthion (paired: exoL and exoR) .....	87
2.5.2 Endocanthion (paired: endoL and endoR).....	88
2.5.3 Palpebrale superius (paired: psupL and psupR) .....	88
2.5.4 Palpebrale inferius (paired: pinfL and pinfR).....	88
2.5.5 Nasion (n) .....	89
2.5.6 Pronasale (prn).....	89
2.5.7 Subnasale (sn).....	89
2.5.8 Alari nasi (paired: alaL and alaR).....	90
2.5.9 Labiale superius (ls) .....	90
2.5.10 Labiale inferius (li) .....	90
2.5.11 Cristae philtri (paired: chpL and chpR) .....	90
2.5.12 Cheilion (paired: chL and chR).....	91
2.5.13 Gnathion (gn).....	91
2.5.14 Otobasion inferius (paired: otoL and otoR).....	91
2.6 Dense surface modelling .....	92
2.7 Statistical analysis.....	100
2.7.1 Face Shape Difference (FSD) .....	100
2.7.2 Reflected Face Shape Difference .....	102
2.7.3 Power calculation .....	102
2.7.4 Statistical tests used with FSD.....	105
2.8 Ethics approval.....	105

2.9 Data protection .....	105
Chapter 3: Validation of stereophotogrammetry and dense surface modelling.....	107
3.1 Introduction .....	107
3.1.1 Validation studies of current stereophotogrammetry cameras.....	108
3.1.2 Studies of accuracy of stereophotogrammetry .....	109
3.1.3 Studies of reproducibility of stereophotogrammetry .....	110
3.1.4 Validation studies of Canfield Vectra CR10 Camera.....	114
3.1.5 Image capture and registration error .....	115
3.1.6 Head position .....	116
3.1.7 Facial expressions .....	117
3.1.8 Subjects with facial dysmorphism.....	118
3.1.9 Landmarking error.....	119
3.1.10 Dense surface modelling error .....	121
3.1.11 Summary.....	122
3.2 Hypotheses .....	123
3.3 Methods.....	123
3.3.1 Subjects and dense surface models.....	123
3.3.2 Image capture and landmarking .....	127
3.3.3 Inter-operator reproducibility error .....	132
3.3.4 Intra-operator reproducibility error .....	136
3.3.5 Calibration error .....	137
3.3.6 Head positioning.....	138
3.3.7 Facial expressions .....	138
3.4 Results .....	138
3.4.1 Subject characteristics.....	138
3.4.2 Inter-operator and intra-operator reproducibility for landmark positions	139
3.4.3 Intra-operator reproducibility in FSD across two DSMs in people with epilepsy .....	141
3.4.4 Overall repeat image reproducibility of FSD in the same DSM.....	142
3.4.5 Camera calibration.....	144
3.4.6 Head positioning error .....	145
3.4.7 Image file size and FSD.....	152
3.4.8 Inter-landmark distances and FSD.....	154
3.4.9 Determination of head position using inter-landmark distances.....	157
3.4.10 Facial expressions.....	160

3.5 Discussion .....	165
3.6 Conclusion.....	170
Chapter 4: Pathogenic structural variants and face shape.....	172
4.1 Introduction .....	172
4.2 Hypotheses .....	173
4.3 Methods.....	173
4.3.1 Recruitment, image capture and face shape analysis.....	173
4.3.2 Classification of brain imaging, intellectual disability and drug history .....	173
4.3.3 Classification of dysmorphism, facial injuries and expression.....	174
4.3.4 Molecular analysis.....	175
4.3.5 Statistics .....	176
4.4 Results .....	176
4.4.1 Subject population .....	176
4.4.2 Face Shape Difference in patients with and without pathogenic SVs in the training set.....	180
4.4.3 Face Shape Difference in patients with and without pathogenic SVs in the validation set.....	181
4.4.4 Face shape characteristics in people with pathogenic structural variants ..	188
4.4.5 Comparison of FSD with inspection of faces by neurologists .....	189
4.4.6 Exclusion of all FSD values that are outliers .....	189
4.4.7 Exclusion of images from children and European sites.....	190
4.4.8 Exclusion of images with facial injury.....	190
4.4.9 Exclusion of images with mouth open .....	191
4.4.10 Exclusion of images in patients with moderate-severe intellectual disability .....	192
4.4.11 Anti-epileptic drug history.....	195
4.4.12 Body mass index (BMI).....	197
4.5 Discussion .....	198
4.6 Conclusion.....	203
Chapter 5: Structural variants analysis.....	204
5.1 Introduction .....	204
5.2 Hypotheses .....	207
5.3 Methods.....	207
5.3.1 Recruitment, image capture and face shape analysis.....	207
5.3.2 Data collection .....	210
5.3.3 Bioinformatics analysis.....	211

5.3.4 Statistical analysis .....	213
5.4 Results .....	213
5.4.1 Pathogenic SV size and gene content .....	213
5.4.2 Gene expression in the embryonic forebrain for genes contained in pathogenic SVs and the relationship to FSD.....	214
5.4.3 Sensitivity analysis performed by method of SV detection.....	216
5.4.4 FSD in patients with any SV over 500kb.....	216
5.5 Discussion .....	222
5.6 Conclusion.....	226
Chapter 6: Face shape and the effect of body mass index and ageing.....	227
6.1 Introduction .....	227
6.2 Hypotheses .....	230
6.3 Methods.....	230
6.3.1 Recruitment, image capture and face shape analysis.....	230
6.3.2 Statistics .....	232
6.4 Results .....	235
6.4.1 Subject population .....	235
6.4.2 Facial shape and association with BMI and body weight.....	236
6.4.3 Graphical representation of facial change with BMI.....	237
6.4.4 Facial shape change with height.....	240
6.4.5 Sensitivity analysis for age .....	240
6.4.6 Adjustment of FSD for BMI .....	241
6.4.7 Longitudinal study of changes in fsd with age.....	243
6.5 Discussion .....	249
6.6 Conclusion.....	255
Chapter 7: Face symmetry in developmental brain lesions .....	256
7.1 Introduction .....	256
7.2 Hypotheses .....	257
7.3 Methods.....	257
7.3.1 Recruitment, image capture and face shape analysis.....	257
7.3.2 Dense surface modelling.....	259
7.3.3 Reflected face shape difference .....	260
7.3.4 Classification of developmental brain lesions .....	260
7.3.5 Statistical analysis .....	261
7.4 Results .....	262

7.4.1 Subject population .....	262
7.4.2 Reflected FSD in people with developmental brain lesions.....	264
7.4.3 Reflected FSD in unilateral and bilateral lesions.....	265
7.4.4 Periorbital, perinasal and lower face regions.....	265
7.4.5 Hippocampal sclerosis and reflected FSD .....	265
7.4.6 Analysis after exclusion of acquired facial deformity.....	267
7.4.7 Reflected FSD and association with age and sex of participants .....	268
7.4.8 Reflected FSD and presence of structural variants.....	269
7.4.9 Reflected FSD and body mass index .....	269
7.4.10 Handedness and reflected FSD.....	270
7.5 Discussion .....	271
7.6 Conclusion.....	276
Chapter 8: Face shape in HNF1B haploinsufficiency .....	277
8.1 Introduction .....	277
8.2 Hypothesis .....	278
8.3 Methods.....	279
8.3.1 Recruitment, image capture and face shape analysis.....	279
8.3.2 Genetic analysis.....	279
8.3.3 Statistical analysis .....	280
8.4 Results .....	281
8.4.1 Subject population .....	281
8.4.2 FSD and reflected FSD in people with HNF1B gene abnormalities.....	282
8.4.3 Visual inspection of mean faces.....	283
8.5 Discussion .....	285
8.6 Conclusion.....	287
Chapter 9: Discussion .....	288
9.1 Stereophotogrammetry .....	288
9.2 Reproducibility studies.....	289
9.3 Dense surface modelling and FSD .....	292
9.4 Reference subjects.....	294
9.5 Dense surface modelling and the relationship to BMI and ageing .....	295
9.6 FSD and genetic disorders including SVs and HNF1B mutations or deletions....	296
9.7 Reflected FSD and cortical malformations including hippocampal sclerosis .....	300
9.8 Limitations of the present study.....	302
9.8.1 Recruitment of reference subjects.....	302

9.8.2 Recruitment of people with epilepsy.....	302
9.8.3 Stereophotogrammetry.....	304
9.8.4 Landmarking .....	304
9.8.5 Dense surface modelling and FSD.....	305
9.9 Conclusions.....	305
9.10 Future directions .....	307
9.10.1 Stereophotogrammetry.....	307
9.10.2 Dense surface modelling and FSD .....	308
9.10.3 Genetic studies .....	309
9.10.4 Imaging studies.....	310
Appendix 1: Image capture protocol.....	312
References .....	318

## LIST OF ABBREVIATIONS

2D	Two dimensional
3D	Three dimensional
aCGH	Array comparative genomic hybridization
Ala L/R	Alari nasi (left/right)
BH	Band heterotopia
BMI	Body mass index
Ch L/R	Cheilion (left/right)
Chp L/R	Cristae philtri (left/right)
CNV	Copy number variant
CT	Computed tomography
DNT	Dysembryoplastic neuroepithelial tumour
DSM	Dense surface model
Endo L/R	Endocanthion (left/right)
Exo L/R	Exocanthion (left/right)
FCD	Focal cortical dysplasia
FISH	Fluorescence in situ hybridization
FSD	Face shape difference
Gn	Gnathion (soft tissue)
<i>HNF1B</i>	Hepatocyte Nuclear Factor-1 Beta
HS	Hippocampal sclerosis
ICC	Intraclass correlation coefficient
Li	Labiale inferius
Ls	Labiale superius
MCD	Undefined malformation of cortical development
MRI	Magnetic resonance imaging
N	Nasion

Oto L/R	Otobasion inferius (left/right)
PC	Principal component
Pinf L/R	Palpebrale inferius (left/right)
Psup L/R	Palpebrale superius (left/right)
Prn	Pronasale
PVH	Periventricular heterotopia
ROC	Receiver operating characteristic
Sn	Subnasale
SNP	Single nucleotide polymorphism
SV	Structural variant
TS	Tuberous sclerosis

## LIST OF FIGURES

Figure 1 A typical phrenology chart.....	22
Figure 2 Craniofacial pattern profile in Sotos syndrome.....	31
Figure 3 Calculation of nasal volume .....	41
Figure 4 Development of the human face .....	48
Figure 5 Genetic syndromes associated with facial dysmorphism and seizures.....	52
Figure 6 Genetic syndromes with seizures by extent of facial dysmorphism.....	53
Figure 7 Study recruitment and groups used to analyse structural variants.....	74
Figure 8 Study recruitment and groups used to analyse body mass index.....	76
Figure 9 Study recruitment and groups used to analyse ageing and asymmetry .....	78
Figure 10 The Canfield Vectra CR10 stereophotogrammetry camera system .....	80
Figure 11 Position used for image capture of a subject's face .....	81
Figure 12 Screenshot from the Mirror Vectra CR10 software program.....	81
Figure 13 Face surface image after image capture.....	83
Figure 14 Face surface image in a third party image viewer .....	84
Figure 15 A typical face surface image with landmarks.....	87
Figure 16 An example of Procrustes analysis.....	93
Figure 17 Thin plate spline.....	94
Figure 18 A wireframe view of the mean face surface in one of the studies. ....	96
Figure 19 Example of the effect of the first two principal components.....	98
Figure 20 Dense surface models of the face .....	99
Figure 21 Cross-section of the face with landmarks.....	121
Figure 22 Inter-operator landmark reproducibility study .....	124
Figure 23 Intra-operator landmark reproducibility study .....	124
Figure 24 Intra-operator landmark reproducibility study and original DSM.....	125
Figure 25 Intra-operator landmark reproducibility study and two DSMs.....	126
Figure 26 Overview of reproducibility studies .....	126
Figure 27 Axes of rotation of the camera .....	128
Figure 28 Raw 3D face images for one participant .....	131
Figure 29 Four inter-landmark distances around the eyes .....	134
Figure 30 Five inter-landmark distances of the nose and mouth regions.....	135
Figure 31 Three inter-landmark distances of facial dimensions.....	136
Figure 32 Images of one control subject with different camera angles.....	148
Figure 33 Examples of face images with high FSD values .....	151
Figure 34 Scatter plot of file size and FSD of a face image .....	153
Figure 35 Relationship between FSD and inter-landmark distance reproducibility ....	155
Figure 36 Distribution of FSD for different camera angles.....	156
Figure 37 Determination of facial angle relative to the camera.....	158
Figure 38 Distribution of FSD with four facial expressions.....	164
Figure 39 Distribution of FSD in patients with pathogenic SVs in training set .....	180
Figure 40 ROC curves of FSD .....	182
Figure 41 Correlation of FSD in two DSMs. ....	184
Figure 42 Distribution of FSD in people with pathogenic SVs in both sets.....	187
Figure 43 Association between FSD and intellectual disability.....	193
Figure 44 Scatter plot of verbal IQ with FSD.....	194

Figure 45 Scatter plot of FSD with BMI .....	197
Figure 46 Study recruitment and groups used to analyse structural variants .....	208
Figure 47 Recruitment of subjects who had the complete genotype data available.....	209
Figure 48 Gene expression figure for a selected gene, AATF .....	212
Figure 49 Scatter plot of number of genes in all pathogenic SVs with FSD.....	214
Figure 50 Scatter plot of number of highly expressed genes and FSD.....	215
Figure 51 Association of number of deletions over 500 kilobases and FSD.....	219
Figure 52 Scatter plot of FSD and the cumulative length of all deletions over 500kb..	220
Figure 53 Distribution of FSD in people with deletions over 500 kilobases .....	221
Figure 54 Study recruitment and groups used to analyse body mass index.....	231
Figure 55 Recruitment of subjects for study of face shape change with age.....	232
Figure 56 Relationship between one principal component and BMI .....	233
Figure 57 The mean face of women and men by BMI .....	240
Figure 58 BMI-adjusted FSD and the relationship to pathogenic SVs.....	242
Figure 59 ROC curve of BMI-adjusted FSD .....	243
Figure 60 Correlation between initial and repeat FSD of whole face.....	245
Figure 61 Correlation between initial and repeat FSD of periorbital region .....	245
Figure 62 Correlation between initial and repeat FSD of perinasal region .....	246
Figure 63 Correlation between initial and repeat FSD of perioral region.....	246
Figure 64 Scatter plot of change in FSD with time interval of images.....	247
Figure 65 Images of subjects with the largest FSD changes.....	249
Figure 66 Recruitment of subjects into group 6.....	258
Figure 67 Distribution of reflected FSD by developmental brain lesion.....	264
Figure 68 Association of face asymmetry and hippocampal sclerosis.....	266
Figure 69 Relationship of FSD with people with <i>HNF1B</i> gene abnormalities.....	282
Figure 70 The mean face of people with <i>HNF1B</i> abnormalities.....	284
Figure 71 Face asymmetry and <i>HNF1B</i> abnormalities .....	284

## LIST OF TABLES

Table 1 Dysmorphological terms and definitions.....	27
Table 2 List of syndromes in Smith’s recognizable patterns of human malformation....	54
Table 3 Summary of characteristics of the different groups used in studies .....	72
Table 4 Summary of the DSMs made for each group of subjects.....	100
Table 5 Studies of accuracy and precision of stereophotogrammetry.....	112
Table 6 Studies of accuracy and precision of stereophotogrammetry.....	113
Table 7 Sequences of images taken per participant.....	130
Table 8 Inter-landmark distances measured for reproducibility error.....	134
Table 9 Age and sex characteristics of the twenty subjects in each set.....	139
Table 10 Intra-operator and inter-operator landmark reproducibility .....	140
Table 11 Reproducibility of inter-landmark distances.....	141
Table 12 Correlation of FSD values in images landmarked twice.....	142
Table 13 Median FSD reproducibility error .....	144
Table 14 Reproducibility of inter-landmark distances with calibration.....	144
Table 15 Reproducibility of inter-landmark distances by camera angles .....	146
Table 16 Median FSD values in seven different camera angles .....	147
Table 17 Grid of reproducibility errors different angles of camera pitch and yaw .....	149
Table 18 Displacement of an inter-landmark distance with camera rotation.....	159
Table 19 Inter-landmark distance reproducibility with facial expressions .....	161
Table 20 Summary of facial expressions and inter-landmark distances .....	162
Table 21 Facial expression and FSD .....	162
Table 22 Summary of all subjects recruited for pathogenic SV study.....	177
Table 23 Detailed characteristics of all patients with pathogenic SVs.....	179
Table 24 DSMs of the 3 different facial regions .....	185
Table 25 Neurology trainees and assessment of facial dysmorphism .....	189
Table 26 FSD values with and without pathogenic SVs.....	190
Table 27 Categories of acquired facial deformities .....	191
Table 28 Summary of sensitivity analyses.....	195
Table 29 Comparison of previous or current anti-epileptic drug use.....	196
Table 30 Sensitivity analysis of subjects within the normal BMI range.....	198
Table 31 Relationship of patients with and without SVs >500kb and FSD.....	217
Table 32 Characteristics of all participants by sex and BMI category.....	236
Table 33 Standard multiple regression analysis for BMI, body weight and FSD.....	237
Table 34 Simple linear regression equations for the effect of BMI .....	241
Table 35 Change in FSD in DSMs between two images for outliers .....	248
Table 36 Details for all participants with developmental brain lesions.....	263
Table 37 Median reflected FSD and people with developmental brain lesions .....	268
Table 38 The relationship of handedness of subjects to reflected FSD. ....	271
Table 39 Details for all subjects with <i>HNF1B</i> abnormalities.....	281
Table 40 Characteristics of subjects with <i>HNF1B</i> gene abnormalities.....	282

## ACKNOWLEDGEMENTS

This thesis was conducted under the supervision of Sanjay Sisodiya and Josemir (Ley) Sander and I am very grateful for their guidance and their patience.

I am most indebted to the hundreds of patients and relatives who very kindly agreed to take part in this study, and to the nurses and clinicians who helped with recruitment and data collection.

Specifically, I would like to thank Emanuele Bartolini, Jan Novy and Lisa Clayton for their assistance in recruiting patients in Chalfont. Carla Marini, Melania Falchi, Raoul Hennekam and Chantal Depondt were very generous hosts for our recruitment of patients in Europe. Dalia Kasperavičiūtė, Claudia Catarino, Mar Matarin, Elizabeth Caruana Galizia, Anna Tostevin, Natasha Schoeler, Andreja Avbersek, Antonietta Coppola and Ana Bartmann were a pleasure to share an office with.

I also thank Raoul Hennekam, Peter Hammond and Mike Suttie who were a source of wisdom and advice about facial shape and dysmorphology. Mike Suttie was always accessible and helpful with any software-related issues that I had during the study.

Isabel Parees, Tabish Saifee, Jasper Morrow and Jan Novy very kindly agreed to analyse images for dysmorphism. Likewise, I thank all 20 control subjects who endured multiple images of their face for reproducibility studies.

I would also like to thank Coralie Bingham, Rhian Clissold and Sian Ellard for enabling the recruitment of people with *HNF1B* haploinsufficiency.

This work was supported by the Wellcome Trust [084730]; UCLH CRDC [F136]; Epilepsy Society; The Freemasons' Grand Charity; The Katy Baggott Foundation; the National Institute for Health Research [08-08-SCC]; Action Medical Research; the Henry Smith Charity; the Fonds National de la Recherche Scientifique [FC 63574/3.4.620.06F to C.D.]; and the Fonds Erasme, Université Libre de Bruxelles.

The development of the ShapeFind software at UCL has been supported by the UK charity NewLife and the US National Institutes of Health.

This work was undertaken at University College London Hospitals and University College London, which received a proportion of funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme.

## AUTHOR'S CONTRIBUTION

All work is by the author with the following contributions by other people:

Chapters 3, 4, 5, 6, 7, 8, 9: Emanuele Bartolini, Jan Novy and Lisa Clayton helped to recruit some patients for image capture. Peter Hammond and Michael Suttie had previously recruited control subjects who were used in the dense surface models as references only in all studies except for Chapter 9, in which they were used for direct comparison. The dense surface models were all created using software created by Peter Hammond and Tim Hutton.

Chapter 4: Peter Hammond provided data on his landmarking for inter-operator reproducibility assessment.

Chapters 5 and 6: Emanuele Bartolini, Carla Marini, Melania Falchi, Renzo Guerrini and Chantal Depondt helped with recruitment of subjects for image capture in European sites. Zoe Fox advised on statistical tests that were appropriate.

Chapter 5: Jan Novy helped to classify facial injury and deformity as an independent observer. Lisa Clayton provided data for some patients on anti-epileptic drug use history.

Chapter 6: Dalia Kasperavičiūtė and Erin Heinzen provided access to raw SNP array data. Jan Novy provided some of the re-analysis of this data using QuantiSNP.

Chapter 8: Sanjay Sisodiya classified brain abnormalities on MRI scans.

Chapter 9: Coralie Bingham, Rhian Clissold and Sian Ellard provided genotype data for all but one of the subjects with *HNF1B* gene abnormalities. Dalia Kasperavičiūtė provided genotype data on the remaining subject, who had also undergone image capture by Peter Hammond.

## PUBLICATION

Part of this thesis has been published as follows:

Chinthapalli K, Bartolini E, Novy J, Suttie M, Marini C, Falchi M, Fox Z, Clayton LM, Sander JW, Guerrini R, Depondt C, Hennekam R, Hammond P, Sisodiya SM. Atypical face shape and genomic structural variants in epilepsy. *Brain*. 2012 Oct;135 (Pt 10):3101-14. Epub 2012 Sep 13. doi: 10.1093/brain/aws232. PubMed PMID: 22975390.

# CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

## 1.1 THE STUDY OF FACE SHAPE

The face is one of the most important visual stimuli in humans and is a prime determinant of perceptions of identity, age, sex, ethnicity, health, sexual attractiveness, mood, and attention (Leopold and Rhodes, 2010). Indeed, natural selection has played an important role in variation of human face shape (Guo et al., 2014). It is unsurprising that humans have analysed facial shape and appearance throughout history and now use advanced imaging for this purpose, as I will outline below.

This introduction is divided into three sections. Firstly, I will discuss the study of face shape from physiognomy to dysmorphology. Secondly, I will review the methods used to measure face shape, including three dimensional (3D) stereophotogrammetry. Lastly, I will explore the relationship between epilepsy, genomics and face shape.

### 1.1.1 ARISTOTLE AND PHYSIOGNOMY

Aristotle was the first known person to document inferences about character and personality from facial appearance, in his treatise, *Physiognomonica*. For example, regarding the nose, he stated:

*A nose broad at the tip means laziness, as witness cattle: but if thick from the tip, it means dullness of sense, as in swine; if the tip is pointed, irascibility, as in dogs; whilst a round, blunt tip indicates pride, as in lions. Men with a nose thin at the tip have the characteristics of birds. When such a nose curves slightly right away from the forehead, it indicates impudence, as in ravens: but when it is strongly aquiline and demarcated from the forehead by a well-defined articulation, it indicates a proud soul, as in the eagle; and when it is hollow, with the part next the forehead rounded and the curve rising upwards, it signifies lasciviousness, as in cocks. (Aristotle, trans. 1913)*

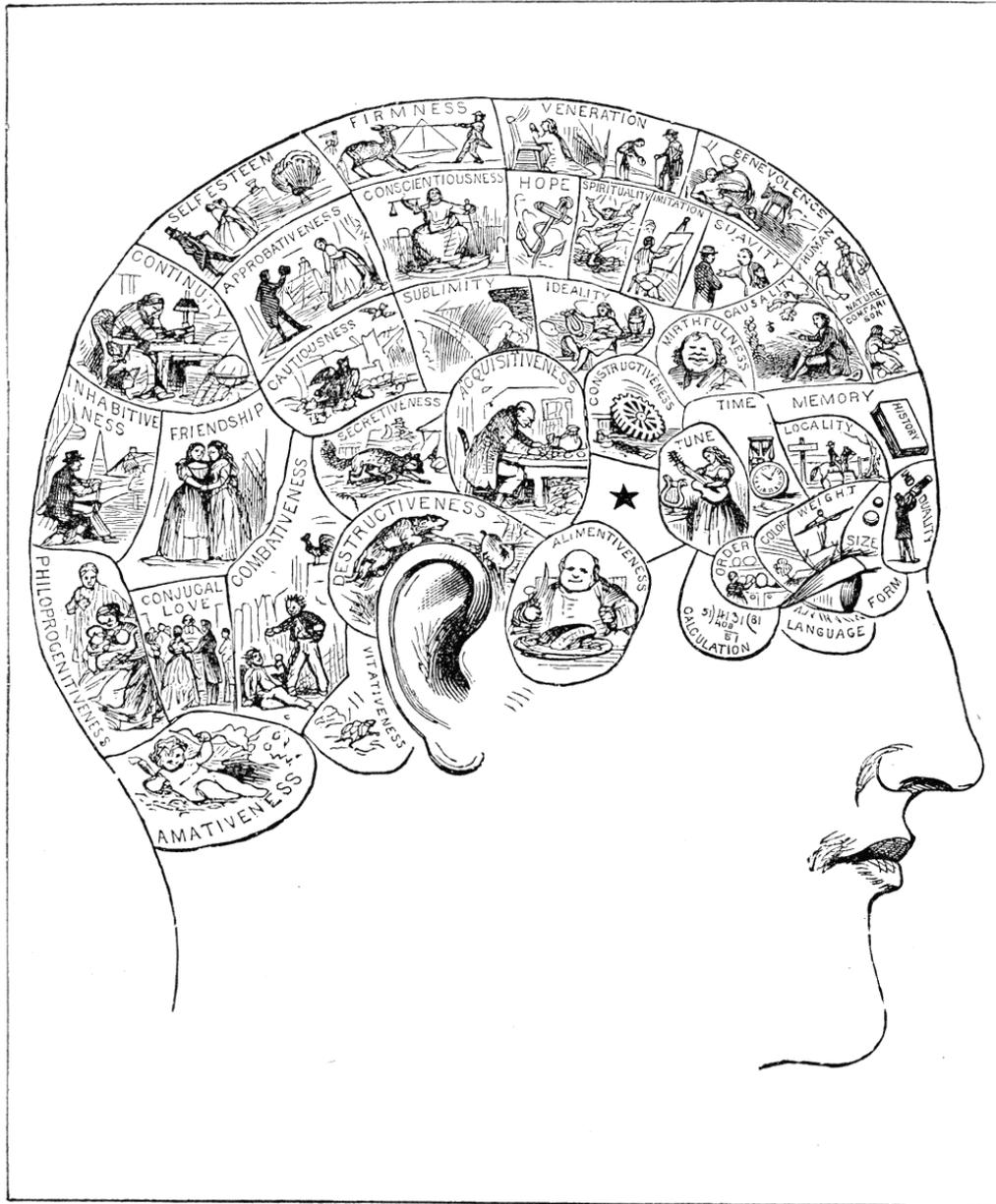
These findings developed into the field of physiognomy (previously termed 'physiognomony') in the Middle Ages when Islamic and European students read Aristotle's work and scholars expanded on his observations of face and body shape (Ziegler, 2007). Ibn Sina, a famous ninth century Persian physician, thought of physiognomy as a practical science alongside medicine, astrology and alchemy. Michael Scot, a 13<sup>th</sup> century Scottish scholar, thought of it as a natural science and suggested that changes in appearance of the face or body occur due to the body's humours. Physiognomers attached special importance to face shape, with one historian stating that the face "being the window to all cerebral powers, was the most individual of all bodily organs. In the face appeared most clearly the differences between individuals of the same species, so that it was the best source of information for casting a physiognomic judgement" (Ziegler, 2007).

Despite the contribution of Scot and other proponents, a glaring weakness of physiognomy was that nobody could scientifically explain the mechanism by which facial (or bodily) shape reflected a person's character. Thus, it waned in popularity during the Renaissance. It was later resurrected by a Swiss priest, Johann Lavater, in the 18<sup>th</sup> century. Lavater thought "The natural and essential bony skull and its carefully measured proportions are the true indicators of the character of man" (Stemmler, 1993).

His publications on physiognomy included illustrations of silhouetted face contours taken in profile view. Without any known empirical evidence, Lavater speculates on the significance of certain features seen in silhouette form. For instance, the angle below the nose is associated with wisdom and the absence of straight lines is associated with a good nature (Stemmler, 1993). Lavater was not taken seriously in scientific circles, but his ideas popularised physiognomy in the following century, and in particular paved the way for phrenology.

### 1.1.2 GALL AND PHRENOLOGY

Franz Joseph Gall, a German physician, was the founder of phrenology, an offshoot of physiognomy. Phrenology was the study of skull shape and its relationship with mental and psychological attributes. Gall believed a neuroanatomical hypothesis that the brain was divided into discrete areas, each responsible for a specific psychological quality. The size of these areas was correlated with the importance of their corresponding psychological quality. Gall then went further and stated that the size of each brain area could be calculated from measurements of the external surface of the skull. His first observation occurred as a schoolchild, when he apparently saw that students with excellent verbal memory tended to have prominent eyes (Simpson, 2005). Therefore, the faculty of memory was located in the inferior frontal part of the brain according to Gall (Eling et al., 2011). He measured plaster casts of live subjects' heads or the skulls of deceased subjects. He studied hundreds of people including psychiatric patients, criminals, artists and others with particular psychological traits (Simpson, 2005).



Phrenological Chart of the Faculties.

**Figure 1** A typical phrenology chart from 1883 showing discrete anatomical parts of the brain and the psychological qualities that they were thought to represent. Representation occurred bilaterally in both cerebral and cerebellar hemispheres (Taken from the public domain).

Gall correctly surmised a few aspects of brain function localisation in his phrenological map of brain areas and their qualities. He predicted that unique human qualities, such as language, calculation and causality were located in the frontal lobes, which were larger than in other animals (Figure 1). Nevertheless, physiologists found that many of Gall's brain areas either performed no apparent function or one completely different to

his theory. For example, physiological experiments showed that the cerebellum was responsible for motor co-ordination and not sexual function, which Gall had proposed. By the 1840s, physiologists also asserted that there was no evidence of a relationship between the external skull surface and the brain shape (Parssinen, 1974).

Phrenology had already been widely disseminated with the aid of Gall's assistant, Spurzheim. In the UK, phrenology societies and journals were set up. A national newspaper said phrenology was believed by "a very large proportion of the shrewdest and best-informed of the middle and labouring classes" and one phrenology publication was said to have the fourth highest all-time readership of any book in the UK (Parssinen, 1974). One of the major legacies of the discredited science of phrenology is its endurance in current popular consciousness. For example, the specialty of dysmorphology was linked unfavourably to phrenology in an international newspaper as recently as November 2015 (The Economist, 2015).

### 1.1.3 LOMBROSO AND EPILEPSY

Within neurology, the development of phrenology by Gall was refined by Cesare Lombroso, another man who was later discredited. Lombroso was an Italian neuropsychiatrist, who developed an interest in criminals and forensic anthropology. He made a link between criminals and people with epilepsy, describing it thus in 1906: "... the idea that the great criminality was a kind of equivalence of epilepsy came through my mind I began to rummage among epileptic skeletons and skulls and I found the same proportions in the median occipital dimple and in the facial asymmetry" (Monaco and Mula, 2011).

In 1871, he had found a dimple in the region of the occipital crest in an autopsy of a bandit and described the presence of such a dimple as being a sign of epilepsy and criminal behaviour. He subsequently analysed the facial appearance of 410 people with epilepsy and stated that a quarter of them had facial features in common with

criminals<sup>32</sup>. Needless to say, there has been no corroboration of this but Lombroso's work perpetuated the stigma of epilepsy.

#### 1.1.4 DYSMORPHOLOGY

Clinical genetics is a small medical specialty, developing from the 1960s onwards, and currently there are 3.2 clinical genetics consultants per million population in the United Kingdom (Federation of the Royal Colleges of Physicians of the UK, 2011), and 3.5 board-certified clinical geneticists per million population in the United States (Cooksey et al., 2006). Dysmorphology, a subspecialty of clinical genetics and paediatrics, is the study of people with congenital malformations and birth defects. The term was coined in 1966 by David Smith, one of the pioneers of the discipline (Smith, 1966). Notably, he was not only interested in craniofacial abnormalities but also in explanations of how they arose. For example, Smith describes cleft lip as a malformation due to defective fusion of the lateral and median nasal processes, occurring prior to 36 days' gestation, and being associated with almost ten times greater concordance in monozygotic twins than in dizygotic twins (Smith, 1966).

For approximately one-third of all clinical geneticists, their main patient group (defined as >40% of all patients seen) comprises those with birth defects or dysmorphology (Cooksey et al., 2006). Traditionally, abnormalities of face structure and shape are under the remit of the dysmorphologist. Genetics and facial shape are closely linked, with up to 40% of genetic disorders showing a craniofacial manifestation (Hart and Hart, 2009). Heritability for human face shape ranges from 0.51 to 1.0 in a review of studies of twins and populations using cephalometry and from 0.61 to 0.87 using anthropometry (Kohn, 1991). A more recent study used a digitizer probe to measure surface points on human skulls from a collection in which familial relationships between the deceased people were known. They found significant heritability for 72% of craniofacial distances measured between skulls (Martínez-Abadías et al., 2009).

These findings are also found in studies of individuals using skull radiographs (also termed 'cephalograms').

Such studies have only assessed the skeletal shape, but others have investigated soft tissue face shape. One group found that the genetic contribution to four facial measurements was 62-71% from assessing 79 pairs of twins (Savoie et al., 1998). Another study looked at 16-year-old children and their parents, finding that average heritability was about 0.7-0.8 for facial measurements but much less for dental measurements (Johannsdottir et al., 2005). It is important to note that many facial similarities are not captured by linear or angular measurements of the face (as used in all of the above studies) but nonetheless are well-recognised from human observation of monozygotic twins, families and ethnicities.

There are three aspects of dysmorphology that I will consider here. Firstly, dysmorphologists may sometimes use quantitative methods, such as anthropometry, but more often rely on recognising a facial 'gestalt' and the constellation of features of a given syndrome. Such recognition requires years of experience with highly specialised training, and it has been suggested that the necessary skills cannot be taught formally (Reardon and Donnai, 2007; Robin, 2011).

Secondly, the terminology used to describe findings may be a hindrance. Descriptions and terms are important in communication of dysmorphic features, be it in clinical case reports or in databases of congenital syndromes. Clinical geneticists use these publications and databases to identify syndromes and the need for standardised nomenclature was stated even in the 1980s (Diliberti, 1988).

Such nomenclature has been slow to develop. As stated by an international panel of senior dysmorphologists in 2009: "The terms that clinicians use to describe a body part have gradually evolved in a haphazard and uncoordinated manner, and have not been critically reviewed" (Allanson et al., 2009a). For example, a finding such as 'low-set

ears' may have different meanings to different physicians. There are at least three anatomical reasons and eighteen corresponding diseases for why ears may appear to be low-set but are not truly low-set (Robinow and Roche, 1973). Other examples are the inability to define intermediate forms of polydactyly in the existing classification, or the presence of three overlapping and imprecise terms for a small eye (Biesecker, 2005).

The aforementioned international expert panel went on to publish a list of standardised definitions of a number of terms used in dysmorphology in 2009 (Allanson et al., 2009b; Biesecker et al., 2009; Carey et al., 2009; Hall et al., 2009; Hennekam et al., 2009; Hunter et al., 2009), which I have summarised in Table 1. For each term, the expert authors classified its definition as 'objective' or 'subjective' or 'both' if both alternatives existed. As an example, 'macrotia' was defined objectively as "median longitudinal ear length greater than 2SD above the mean and median ear width greater than 2SD above the mean" and was also defined subjectively as "apparently (sic) increase in length and width of the pinna" (Hunter et al., 2009). Some of these objective definitions rely on available normative values, notably from tables of face shape measurements using direct anthropometry (Farkas, 1994; Feingold and Bossert, 1974; Hall et al., 2007; Merlob et al., 1984). Even in this up-to-date list of dysmorphological features, 62% (244 of 396) of all terms only have subjective definitions. Also, 358 recognised synonyms exist for the standardised definitions adding to the complexity of the nomenclature.

Body area	Nos of terms	Type of definition used			Nos of synonyms	Examples of subjective terms
		Objective	Subjective	Both		
Cranium	10	0	6 (60%)	4	7	Plagiocephaly
Scalp hair	9	2	5 (56%)	2	5	Low posterior hairline
Face	9	0	5 (56%)	4	3	Triangular face
Forehead	12	0	9 (75%)	3	9	Frontal bossing
Maxilla, midface	12	0	12 (100%)	0	17	Midface retrusion
Mandible	6	1	3 (50%)	2		Micrognathia
Chin	8	0	8 (100%)	0	3	Chin dimple
Neck	5	0	5 (100%)	0	4	Neck webbing
Periorbital	39	4	26 (67%)	9	26	Epicanthus
Ear	73	33	37 (51%)	3	70	2 <sup>nd</sup> degree microtia
Nose	41	9	27 (66%)	5	46	Bifid nose
Philtrum	10	2	4 (40%)	4	6	Tented philtrum
Lips	15	3	8 (53%)	4	17	Lip freckle
Mouth	9	2	5 (56%)	2	14	Upturned corners
Oral cavity	36	11	20 (56%)	5	24	Small tongue
Hands, feet	83	23	48 (58%)	12	97	Pes cavus
Creases	7	3	4 (57%)	0	3	Absent palmar crease
Nails	12	0	12 (100%)	0	7	Small nail
<b>All areas</b>	<b>396</b>	<b>93</b>	<b>244 (62%)</b>	<b>59</b>	<b>358</b>	

**Table 1** Dymorphological terms, the objectivity of their definitions, and their synonyms. All of the terms used in a list of standardised definitions of dymorphic features are included according to body area. Each of the 396 terms was defined by the expert panel using what they called 'Objective' criteria, 'Subjective' criteria or 'Both' objective and subjective criteria. Examples of the subjectively defined terms are in the final column. As well as 244 (62%) dymorphological terms only being defined subjectively, there also exist 358 synonyms, adding to the complexity of the nomenclature. The list of terms can be found in a series of six review articles (Allanson et al., 2009b; Biesecker et al., 2009; Carey et al., 2009; Hall et al., 2009; Hennekam et al., 2009; Hunter et al., 2009).

Thirdly, other physicians, especially in adult medicine and general practice, receive little training in dymorphology and clinical genetics, and are less familiar with these fields (Taylor et al., 2006; Watson et al., 1999). Data from a 2003 survey of UK general practice and specialty trainees revealed that 85% of trainee general practitioners and 46% of specialty trainees had no postgraduate genetics training (Burke et al., 2006). Of note, 63% of neurology trainees nevertheless thought that genetics was an element of a case at least once a week. In a more recent 2012 survey of family physicians in the USA and Canada, 54% of respondents said that they were not knowledgeable about genetic testing (Mainous et al., 2013). Also in the USA, surveys have reported that 26% of medical students in clinical training have formal teaching in genetics (Wolyniak et al.,

2015). Similarly, 22% of allied health professionals reported that they had at least 'satisfactory' training in genetics (Christianson et al., 2005). About two-thirds of neurologists and psychiatrists in the USA do not feel confident about genetic testing and 49% of neurologists do not have access to a geneticist (Salm et al., 2014).

On the other hand, only a minority of clinical geneticists specialise in adults – 7% according to one US study (Cooksey et al., 2006). There is concern that a large unidentified population of adults with dysmorphic features and genetic syndromes remain undiagnosed by their physicians, compounded by the fact that multiple physicians may be involved in treatment of the different manifestations of a patient's syndrome (Taylor et al., 2006; Williams, 2007). In one study, the referral of adults to an adult clinical genetics service resulted in 50% of them having their overall diagnosis changed, often on the basis of clinical examination alone (Maves et al., 2007).

In conclusion, dysmorphology is practised by only a few clinicians and it is not necessarily objective or consistent between such practitioners. Many adults with genetic syndromes or dysmorphism may not be identified in the course of their clinical care.

By contrast, there have been significant advances in imaging-based analysis of facial shape, particularly using 3D stereophotogrammetry, and these techniques may be useful in clinical practice.

## 1.2 THE STUDY OF FACE SHAPE MEASUREMENT

### 1.2.1 DIRECT ANTHROPOMETRY

Anthropometry is the study of measurement of humans. Measurements can be made directly on a subject or indirectly from images of the subject. Modern facial anthropometry developed as a method of identification of criminals in late nineteenth century Paris. Alphonse Bertillon devised a system to identify people based on ten

measurements. These included head length, facial width and ear length, but the scope for error in measurements led to its demise in forensics use in favour of fingerprinting (Maguire, 2009).

Craniofacial anthropometry for clinical use was pioneered by Leslie Farkas, a Hungarian plastic surgeon in the late twentieth century. He advocated precise measurement of the human face by using sliding callipers, spreading callipers, angle meters and measuring tape. He also placed emphasis on soft tissue facial 'landmarks' that could be reliably identified in different subjects.

Landmarks are specific anthropometric points on the face surface that are reproducible and precisely defined. Once identified, distances and angles may be calculated between them. Farkas described 47 landmarks on the head and face (Farkas, 1994) and many of these were originally identified by artists during and after the Renaissance (Farkas et al., 1985). As such, many of them can be identified visually either by colour (e.g. lip contour) or shape (e.g. eyelid contour). A remaining few landmarks are derived from cephalometric points and traditionally were identified using palpation, such as soft tissue gnathion: Gnathion is the midpoint of the lower border of the mandible and soft tissue gnathion is the point anterior to that on the skin surface. Others, such as opisthocranion are on the posterior aspect of the head and therefore not facial landmarks.

Most importantly, Farkas used measurements between landmarks in large groups of subjects of different age, sex and ethnicity, to create tables of normative values (Farkas, 1994).

The normative data could be used for clinical measurements. For example, Farkas and colleagues used craniofacial anthropometry in people with Down syndrome to show that 37% of measurements were 'subnormal' or 'supranormal' – i.e. outside normal limits of two standard deviations from the same measurement in the mean age-

matched, sex-matched control face. Subnormal measurements included head circumference in 71%, head length in 66%, midface depth in 63%, and ear length in 72% of people with Down syndrome (Farkas et al., 2001). Farkas took this approach a step further. A list of those measurements that differ most from the normal value were compiled. Then the z-scores of deviation from the mean were plotted on a single axis for each of these measurements. The resulting graph displays the z-scores and the pattern was described as a 'craniofacial pattern profile' (Figure 2). Different subjects can be plotted on the same graph to assess similarity in facial appearance. This method was used for Down syndrome, in which craniofacial pattern profiling had a sensitivity of 97% (Allanson et al., 1993).

***Figure 2 Craniofacial pattern profile for two patients with Sotos syndrome. Not shown as unable to confirm copyright clearance.***

Craniofacial anthropometry is not routinely used in clinical practice despite the above benefits. The use of such instruments is time-consuming and it may be difficult for subjects, such as children, to co-operate with the examination. Farkas describes the need for exact positioning of the subject, with some measurements best taken with the subject supine and others taken in the 'standard' head position (Farkas, 1994). He further notes that an assistant is sometimes needed to ensure the correct position.

Another disadvantage is that measurements can only be taken with the subject present. This may result in inadequately trained individuals taking the measurements, and no opportunity to rectify any errors afterwards (Allanson, 1997). One study found that in 0.5% of measurements by experienced researchers, the value was recorded wrongly, introducing a further error into calculations (Jamison and Ward, 1993).

At present, no study has used an automated form of direct anthropometry. In other industries, co-ordinate measuring machines are used for precise surface measurements in manufacturing but these are unacceptably slow for clinical use. For the human face, co-ordinate measuring machines have only been used for facial casts (Ayoub et al., 2003; Khambay et al., 2008; Lincoln et al., 2015). One related technique, which has not been widely used, is to use a handheld digitiser to trace the contour of a surface or record specific points in 3D space. One centre has used the Microscribe digitiser (Solution Technologies Inc; Oella, MD, USA) to assess 50 soft tissue landmarks on the face (Ferrario et al., 2004). The subject's head had to be fixed in a frame and collection of landmarks, which were already pre-marked on the skin, took approximately one minute by an operator. Accuracy was within 1.1mm of direct anthropometry for eight out of eight distances, with a mean error of 0.22mm (Ferrario et al., 1998).

### 1.2.2 INDIRECT ANTHROPOMETRY (TWO DIMENSIONAL)

Indirect anthropometry is the use of images to perform anthropometric measurements. Two-dimensional (2D) imaging of the face can be in the form of photographs (photogrammetry) or radiographs (cephalometry).

The term 'photogrammetry' was coined in 1867 by Albrecht Meydenbauer, a German surveyor (Albertz, 2007). He described the use of a camera with crosshairs, at a known distance from the target, to provide measurements of the target just from the camera images. Since then, his technique has been successfully used in architecture and mapping. Photogrammetry of the face has been performed for research since 1940

(Beard and Burke, 1967) and unlike standard anthropometry, it allows for a permanent record of facial shape to be stored as two-dimensional images that can be reanalysed if needed. Usually at least two photographs are needed: the first is a frontal view, and the second is a lateral or profile view. An oblique view may also sometimes be taken. A number of shortcomings of photogrammetry have been described.

Farkas found that there was a significant systematic error when using photogrammetry (Farkas, 1994). Three measurements were consistently longer: facial width by an average of 3.6mm, nasal width by 2.4mm, and lower facial width by 21.6mm. Eleven measurements were consistently shorter and another fifteen showed mixed differences. Farkas had two main explanations for this. Firstly, landmarks are not always visible or clear enough on a photograph, even when marked on the subject's face beforehand. This may be due to obscuration by other parts of the subject's face or hair. Secondly, photographic distortion occurs. Diliberti and Olson identified 12 sources of error when a three-dimensional (3D) object is photographed on to two-dimensional film (Diliberti and Olson, 1991). One is projection error due to the short distance between the camera and the subject. For instance, in frontal views, the distance between the tip of the nose and the base of the nose was 46% shorter on photogrammetry compared to direct anthropometry, whereas the distance from the root of the nose to the lips was much more concordant (Farkas, 1994). The measurement error is exaggerated in this case because it is perpendicular to the focusing plane. Poor lighting, poor focusing and poor head positioning also introduce errors. One study, of 14 children and adults with 22q11 microdeletion syndrome, found that 5 out of 14 different facial measurements differed by at least 10% between direct anthropometry and photogrammetry (Guyot et al., 2003). Four of these measurements were taken from frontal photographs in which different landmarks may be at different depths relative to the measurement plane. There is less variability in depth for profile photographs.

A software program has also been developed to analyse facial features from a single colour photograph of a subject and compare it to faces in a database of dysmorphology syndromes. Two case reports of its success in retrospectively detecting facial features of two genetic disorders have been published so far (Gripp et al., 2016).

Cephalometry uses multiple radiographs of the skull and facial skeleton. Measurements of various bony landmarks are taken and compared to normative values. The technique involves exposing subjects to ionising radiation, and does not utilise any soft tissue landmarks, thereby disregarding much information about facial appearance (Allanson, 1997).

### 1.2.3 INDIRECT ANTHROPOMETRY (THREE-DIMENSIONAL)

Over the last two decades, more advanced imaging methods have been developed that can capture and store the three-dimensional (3D) face surface. Three methods capture only the exterior face surface: moiré photography, laser scanning and stereophotogrammetry (see Section 1.2.5). Two methods also capture the underlying craniofacial skeleton: computed tomography (CT) and magnetic resonance imaging (MRI).

Moiré photography of the face was first described in the 1970s. A camera is placed behind a grid and thus projects a moiré pattern of alternating light and dark fringes on the surface of the target object. The fringes are distorted by curvature of the object and resemble contour lines. The surface with the superimposed bands can be captured in a photograph. In one study of face shape, there was a 10% error in length calculations (Madden and Karlan, 1979). Also, to calculate fringe spacing, faces had to be maintained very close to the camera with the nose less than 1mm away. A study using facial soft tissue landmarks also found that accuracy and precision were poor (Douglas, 1986). One group used moiré photography to measure lip morphology in infants and control adult faces (Kawai et al., 1990a, 1990b). However, a frame was used to fix the

head in position along with a long shutter speed of 250ms. Major disadvantages of the technique are limited field of view, poor accuracy and a cumbersome setup.

More recently, a team have compared moiré photography, using a standard projector and personal computer, with stereophotogrammetry (Artopoulos et al., 2014). They found no significant differences between the two with a mean surface error of 0.14mm for moiré photography and 0.15mm for the Di3D stereophotogrammetry system (Dimensional Imaging, Glasgow, UK). However, the study is of limited relevance because the authors used polymer face models with a reflective coating and only analysed the middle third of the face. Moiré photography from one source could not capture the ears or jawline. Also both systems had slow shutter speeds (0.07s for moiré photography and 0.05s for the Di3D system) compared to other stereophotogrammetry devices.

Laser scanning uses a laser and a sensor placed at a known angle and distance from the laser. A spot or stripe (also called 'slit') is projected on to the target surface and its scattered reflection is detected by the sensor. Triangulation is used to calculate the 3D co-ordinates of the surface point. Typically, the laser beam must move across the face in sequential horizontal or vertical stripes. Scanners may be portable. Accuracy has been reported to be  $1.9 \pm 0.8$ mm for face shape measurements using a plaster model (Kau et al., 2007). The principal disadvantage is that scan acquisition times are at least 0.3-7 seconds for the face (Kau et al., 2005), which means accuracy may be reduced due to movement in a living subject. One review describes subject discomfort during laser scanning (Riphagen et al., 2008) and the eyes must be closed with some lasers, thus limiting capture of periocular shape, including the palpebral fissure. As the laser beam originates from one static source, there may be limited coverage of sharply contoured face regions (such as the eyes, nose and lips) causing artefacts. Another disadvantage is

that there is no surface colour information, unlike in stereophotogrammetry, meaning that some facial landmarks may not be identified as accurately.

#### 1.2.4 COMPUTED TOMOGRAPHY AND MAGNETIC RESONANCE IMAGING

Clinical CT scanners use X-rays to construct face structure using stacked slices, usually 5mm apart. An advantage is that detailed information about bony structures underlying the face surface is also known. This allows bony landmarks to be used, which should be less affected by movement or facial expression than soft tissue landmarks. Facial bone structure is also known to be affected in some conditions affecting face shape, such as fetal alcohol syndrome (Sant'Anna and Tosello, 2006). In one study comparing 3D CT to stereophotogrammetry in 70 people, CT could only identify half of all (14 of 28) facial landmarks used (Metzger et al., 2013).

MRI scans of the head can also generate 3D images of the face and underlying structures. They have been used for assessing face shape in children in three studies by one group (Chakravarty et al., 2011; Liu et al., 2012; Marečková et al., 2015). MRI has the potential advantage of capturing detailed brain structure as well as craniofacial soft tissue and skeletal shape. In one study, face shape, particularly jaw length, was found to explain 8% of brain size variance in 12-18-year-old French Canadian girls (Marečková et al., 2015). Brain images were not analysed in the same 3D image as craniofacial landmarks. Instead the brain outline was extracted from the raw MRI images and then brain volume was calculated (Smith et al., 2002), so more sophisticated analysis of the relationship between brain and face shape was not possible.

The main disadvantage of CT scans is that they expose a subject to ionizing radiation. MRI scanners use no ionizing radiation but generate a strong magnetic field and cannot be used in people with some electronic medical devices (e.g. vagal nerve stimulators, pacemakers) or metal implants. The presence of metal, such as dental crowns, may create artefacts on CT images. Both types of scan are relatively contraindicated in

pregnant women. Size and weight restrictions apply to subjects as they must be able to lie on a gantry within a tunnel-shaped scanner. As well as subjects transferring themselves on to a gantry, they must be able to lie still for at least 10 seconds for a CT scan and over a minute for an MRI scan.

Lying supine introduces another error. Facial structures, including soft tissue and the mandible, are known to be distorted by the force of gravity in different postures (Kau et al., 2007). Facial landmarks moved by 1-4mm in the sagittal and in the frontal planes in all subjects lying supine in one study (O'Boyle et al., 1996).

Scanners are very expensive, requiring dedicated staff, and are not portable. The resulting images do not capture any surface colour, which may be important for identifying certain facial landmarks. MRI images may not capture all of the facial surface: in one study, the authors inferred the location of missing facial landmarks using 2D photographs (Liu et al., 2012).

### 1.2.5 STEREOPHOTOGRAMMETRY

One powerful method for investigating facial shape is stereophotogrammetry, an extension of photogrammetry. Stereophotogrammetry is a specific form of photogrammetry that uses at least two or more photographs to determine 3D measurements and co-ordinates.

In the 1930s, clinicians applied cartographic principles to create contour maps of the human face using multiple cameras and plotting machinery (Beard and Burke, 1967). The time, effort and cost needed to produce such contour maps prevented them being used clinically. In 1989, Thomas and colleagues described the use of three 35mm cameras to simultaneously acquire images of a subject's face (Thomas et al., 1989). Landmarks were marked after the images were digitised and then computer software created a 3D map of the landmark co-ordinates. In the last twenty years, at least three

manufacturers have developed commercial stereophotogrammetry systems for the human face (Tzou et al., 2014). The technique has been used in orthodontics (Lane and Harrell, 2008), plastic surgery (Honrado and Larrabee, 2004), forensic science (Evison et al., 2010) and dysmorphology (Hammond et al., 2004). It has also been used to study facial attractiveness (Galantucci et al., 2014) and ethnic variation in face shape (Kau et al., 2010).

Stereophotogrammetry uses at least two cameras to take images simultaneously with subsequent image processing by computer software. Accuracy is reported to be within 0.5mm for all major commercial systems (Riphagen et al., 2008; Tzou et al., 2014).

Image acquisition time is shorter than with other methods and often under 4 milliseconds (Lane and Harrell, 2008). Cameras are portable, emit no laser or ionising radiation and collect surface texture and colour information.

It is important to maintain a fixed, known spatial arrangement between cameras or else accuracy is compromised. In commercial systems, up to six cameras are usually housed in a casing to fix their positions. In addition, before each use, most camera systems must be calibrated to adjust for any small changes in the positions and angles of all cameras.

A stereophotogrammetry device also has computing hardware to create a 3D image and the process has been described in detail (Lane and Harrell, 2008). The image is comprised of two parts: a set of 3D surface co-ordinates describing the geometry of the surface, and a colour photograph of the skin and texture. A computer program combines these to create the final 3D image. The set of 3D co-ordinates may be referred to as a 'point cloud' and comprises tens of thousands of points. Older devices may generate multiple sets of 3D surface co-ordinates and 'stitch' them together, whereas newer systems generate one 3D surface immediately, reducing errors. The colour photograph is then projected on to the 3D surface in an image viewer software

program to allow facial features and landmarks to be more easily identified. However, it should be noted that the colour photograph itself is not used in face shape analysis.

Passive stereophotogrammetry uses natural facial features to identify common points between each camera's image. The method depends on background lighting and may not work for uniform regions of colour and texture. Active stereophotogrammetry combines light projection with stereophotogrammetry. A random light pattern is 'actively' projected on to the face at the exact time of image capture. Projecting light makes the camera less dependent on background lighting conditions and a random pattern is used to help detect facial contours, since uniform areas of colour and texture are the hardest to capture accurately. The projected light pattern is not present in the final generated 3D surface. Note that in practice, most commercial cameras, including the one used for my work, apply passive stereophotogrammetry but also use a photographic 'flash' at the time of image capture. Thus passive systems can also reduce the effect of background light variation on image quality.

A related technique to active stereophotogrammetry is that of structured light projections. Here, a specific non-random light pattern is projected on to the face. The light pattern is deformed by the face contour and captured by one or more cameras. The deformed light pattern itself is used to identify the same point on a face surface in each camera. One widely used structured light system (Axis Three, Miami, FL, USA) was compared to three stereophotogrammetry devices and found to have a slower image capture time of 1-2s and an accuracy of 0.5mm compared to 0.1-0.3mm for stereophotogrammetry (Tzou et al., 2014).

The accuracy of stereophotogrammetry for face images has been assessed in 13 published studies (see Table 5 in Chapter 3), which are discussed in detail in Chapter 3. Accuracy was assessed in most studies by comparing the distance between two face surface landmarks using stereophotogrammetry with the 'gold standard' direct

anthropometric measurement. The mean error from eleven of the studies was 0.52mm and another two reported the error as being less than 1mm without providing exact values. Precision or reproducibility error in seven studies was approximately 0.8mm but varied from 0.06mm in automated repeat imaging of mannequin heads, to 2.8mm for distances on infant scalps between different operators.

## 1.2.6 METHODS OF ANALYSING 3D FACE SHAPES

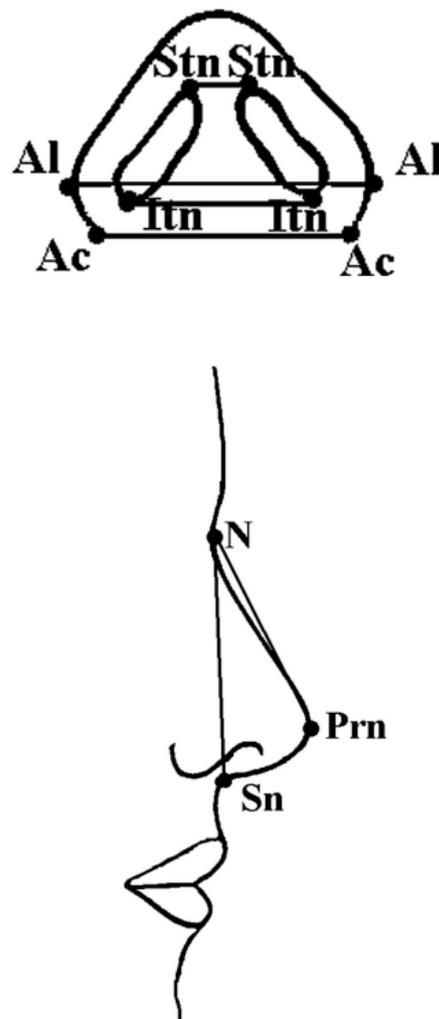
### 1.2.6.1 SHAPE ANALYSIS USING DIRECT ANTHROPOMETRY

Direct anthropometry and indirect anthropometry, including stereophotogrammetry, allow the creation of a 3D co-ordinate map of the face surface. In the instance of direct anthropometry, the 47 classical landmarks described by Farkas could be used as the point co-ordinates. These points were used to measure 103 linear and arc distances, and 29 angles (Farkas and Deutsch, 1996). A three-dimensional map could theoretically be constructed using trigonometry from such measurements. No such map has been created or used in clinical research but combinations of linear and arc measurements have been analysed together.

Farkas had a database of normative distributions for each of the distances between facial landmarks and for each subject, he could create z-scores, which were the number of standard deviations of a measurement from the age-matched, sex-matched mean measurement (Farkas, 1994). As aforementioned, the distribution of z-scores for different facial measurements can be displayed graphically as a craniofacial pattern profile (Figure 2). Building upon this, Farkas and colleagues developed the 'craniofacial variability index', which measured the 'degree of facial disharmony' in subjects (Ward et al., 1998). The z-scores of 16 distances throughout the face are found and then the standard deviation of these z-scores is calculated to give the index. A high craniofacial variability index was found to correlate with subjective impression of dysmorphism in a range of syndromes. However, small numbers of patients with syndromes were

included – no more than four for any syndrome – and all of the syndromes, such as Crouzon syndrome, were chosen because they clearly affect craniofacial features.

Landmarks have been used to derive volumes. An Italian study conducted crude estimates of lip and nose volume from 3D images of people with Down syndrome (Ferrario et al., 2004). This was done by assuming the lip or nose comprised tetrahedra defined by four landmarks each. For example, the nose consisted of one tetrahedron using both alar crests, pronasale and subnasale, and another tetrahedron using the same landmarks except with nasion instead of subnasale (Figure 3).



**Figure 3** Nasal volume can be calculated using five landmarks by approximating the nose to two tetrahedra. The nasal landmarks used are shown in an inferior view and profile view. The first tetrahedron has a base formed from both alar crests (Ac) and pronasale (Prn)

*with the apex formed by nasion (N). The second tetrahedron has the same base but uses subnasale (Sn) as the apex. Figure used with permission from Ferrario et al, 2004.*

In a more mathematical approach, one group in Dublin used only five landmarks of the face (nasion, subnasale, gnathion, left and right tragions) and nine distances between the landmarks to compare face shape in patients with schizophrenia and people without (Hennessy et al., 2004). They used direct calliper measurements for all nine distances and used these to derive 3D point co-ordinates for all five landmarks, which they thought of as the face shape. The authors termed their next step 'geometric morphometrics' (Hennessy and Stringer, 2002). In this, they used generalised Procrustes analysis to align the 3D face shapes together. Then a set of eight principal components was used to describe the variations seen between patients and control subjects, with just three principal components being able to distinguish the two groups of subjects. These processes are also described later. The authors noted that their 3D analysis allowed them to identify subtle differences in face shape that were missed on an earlier analysis, by the same group, based solely on landmarks (Lane et al., 1997). One such difference was of a lengthened lower mid-face height in patients. Thus even a limited analysis using five surface points was more discriminating in some areas than using inter-landmark distances alone.

#### 1.2.6.2 SYMMETRY ANALYSIS USING EARLY STEREOGRAMMETRY

Stereogrammetry also allows for classical anthropometric measurements to be found with high reproducibility and for more complex analysis. Normative data on inter-landmark distances has been created for Italian children using stereogrammetry instead of direct anthropometry (Ferrario et al., 1999). High reproducibility has also been shown in studies of subjects with facial dysmorphism (Aldridge et al., 2005; Heike et al., 2009). The technique also has the potential to analyse facial structure in much more detail than direct anthropometry. Using the first stereogrammetry devices, researchers were able to construct precise lines

between landmarks and then reference planes on digital images. These annotations were used to quantify facial asymmetry and show it is significantly greater in those people with operated cleft palates (Ras et al., 1995). Another group studied facial asymmetry using distances and angles between landmarks, and compared a control population with people affected by one of eighteen syndromes (Shaner et al., 2000). They could find no significant difference in asymmetry and they note a number of limitations. The choice of particular landmarks means that only those landmarks are assessed for asymmetry and not other parts of the face. Also, asymmetry of one landmark may have been compensated for by asymmetry in the other chosen landmark too, as was the case for tragion and subnasale.

#### 1.2.6.3 SHAPE ANALYSIS USING LASER SCANNING

The aforementioned group in Dublin applied geometric morphometrics to analyse sexual dimorphism using laser scanning, showing statistical differences between male and female face shape (Hennessy et al., 2002). Laser scanning, landmarking of 65 points and generalised Procrustes analysis was used by another group to study schizophrenia and they found that the face was vertically elongated in patients (Buckley et al., 2005). These findings should be applicable to stereophotogrammetry too.

Laser scanning was also used by Paternoster and colleagues to collect face images in over 3800 subjects as part of the Avon Longitudinal Study of Parents and Children, for whom genetic data had already been collected (Paternoster et al., 2012). They then carried out a genome-wide association study for 54 landmark distances in the face. One distance between nasion and menton was significantly associated with a SNP in an intron of the *PAX3* gene and this association was also successfully replicated. The authors found that 1.3% of the variation in nasion-menton distance was explained by this polymorphism. There are three factors complicating interpretation of this finding. Menton is not a face surface landmark but rather a derived point from calculations

using other landmarks. No other measurement using menton and no other measurement along the same axis as nasion-menton was found to be associated with any SNP. Thirdly, the *PAX3* gene is implicated in Waardenburg syndrome, a syndrome characterised by telecanthus and a broad nose, but these features are not associated with variation in the nasion-menton distance. On the other hand, a similar genome-wide association study using upper face landmarks from MRI imaging data as well as 2D photography, also found an association between the *PAX3* gene locus and distances between the endocanthion and nasion (Liu et al., 2012).

#### 1.2.6.4 MATRIX ANALYSIS USING STEREOPHOTOGRAMMETRY

Another landmark-based method is the Euclidean distance matrix analysis (Lele and Richtsmeier, 1995). This method involves the creation of a matrix of all distances between all landmarks; for example, with 22 landmarks, there would be  $(22 \times 21 / 2) = 231$  inter-landmark distances. The matrices between different subjects can be compared mathematically without the need to adjust for face size or orientation. This approach has been used in conjunction with standard inter-landmark distance comparisons in one study with conflicting results (Weinberg et al., 2008). Standard inter-landmark distances identified a different pattern of face shape differences from matrix analysis in two groups, controls and relatives of people with cleft lip/palate. Furthermore, the matrix analysis results only appeared to be significant at  $p < 0.1$ , instead of  $p < 0.05$  used for inter-landmark distances. This would suggest matrix analysis offers little improvement over inter-landmark distance comparison. Others have also noted that the technique only assesses distances but not the underlying facial shape, and visualising specific differences in face shape is difficult (Hallgrímsson et al., 2015).

#### 1.2.6.5 CONCLUSION

Even with all of these complex analyses, only a few facial landmarks were studied. This means that the data of the other thousands or tens of thousands of surface point co-ordinates with 3D imaging is discarded. As one review states “Landmarks do not contain information on the spaces, curves, or surfaces between them” (Richtsmeier et al., 2002).

Dense surface modelling is one method that overcomes many of the limitations of landmark-based anthropometry and is only possible with 3D imaging techniques.

#### 1.2.7 DENSE SURFACE MODELLING USING STEREOGRAMMETRY

Dense surface models (DSMs) of the face were developed by Tim Hutton and Peter Hammond (Hutton, 2004; Hutton et al., 2001). The method allows complex shape surfaces to be compared with each other in their entirety. The process begins with landmark-based anthropometry. A sparse set of landmarks is manually identified by an operator on a 3D face image. Each landmark is said to have correspondence between different images because it identifies the same anatomical point. The set of landmarks is then used to co-register face surfaces together. Once registered, a dense correspondence is created from the previous sparse (landmark-based) correspondence. The dense correspondence uses all of the surface point co-ordinates of the specified region, which may be the whole face or just one part of the face. Principal component analysis is used to reduce the dataset of approximately 20,000 surface points to 30-200 principal components, which describe almost all of the face shape variation. The process is described in more detail in Chapter 2.

Dense surface modelling was first applied to images captured using 3dMD DSP400 and 3dMD MU2 camera systems (Hammond et al., 2004). It has also been used with images captured by the Canfield Vectra CR-10 system (Cox-Brinkman et al., 2007; P. Hammond et al., 2008). It has been successfully used to discriminate face images of control

subjects from those with 22q11 deletion syndrome, Cornelia de Lange syndrome, Fabry disease, fetal alcohol syndrome, fibrodysplasia ossificans progressiva, fragile X syndrome, Noonan syndrome, Prader-Willi syndrome, Smith-Magenis syndrome, Williams syndrome and Wolf-Hirschhorn syndrome (Bhuiyan et al., 2006; Cox-Brinkman et al., 2007; de Souza et al., 2013; Hammond et al., 2012a, 2012b, 2005, 2004; Heulens et al., 2013; Suttie et al., 2013). DSM-based analysis identified an unrecognised facial phenotype in mild Wolf-Hirschhorn syndrome and Bardet-Biedl syndrome (Hammond et al., 2012a; Tobin et al., 2008). Another DSM-based analysis could predict the facial shape of people with genetic deletions based on the shape of the faces of people with corresponding duplications at the same locus (Hammond et al., 2014).

The Dublin group, mentioned earlier in Section 1.2.6.1, later developed their technique to incorporate further surface points after manual landmarking, which they call 'pseudolandmarks'. Up to 1700 pseudolandmarks are incorporated into an image by using geometric morphometrics combined with dense correspondence creation, which is done using the method developed by Hutton. Using this they found an increased facial width and altered facial height in people with schizophrenia or bipolar disorder (Hennessy et al., 2010, 2007).

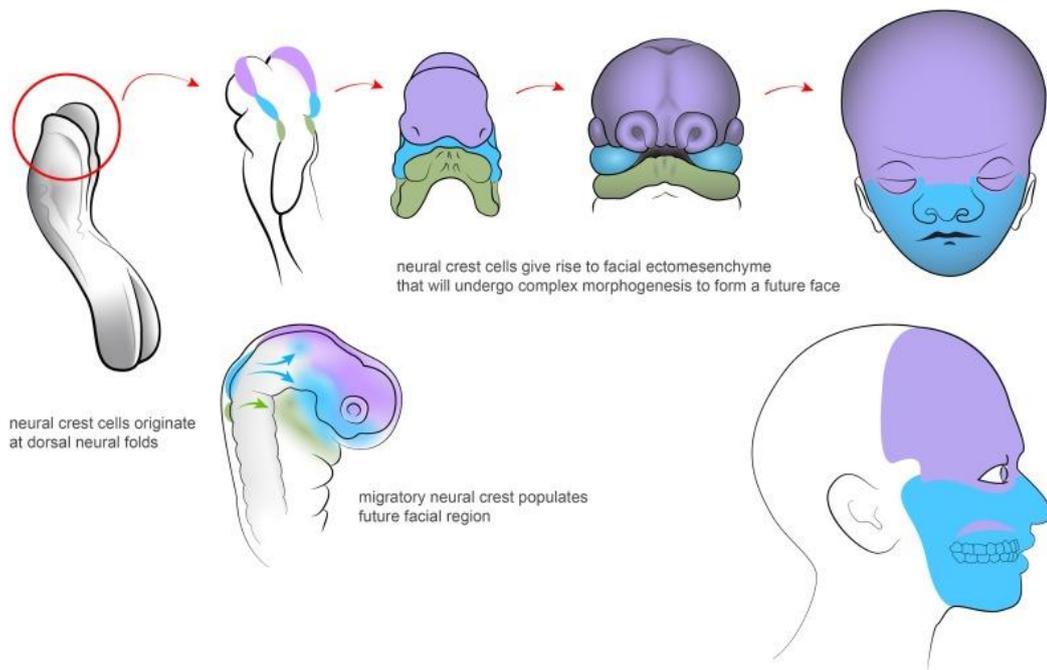
Another more recent study also combined 3D stereophotogrammetry with dense surface modelling to assess the effect of candidate genes on facial shape in three populations of mixed ancestry (Claes et al., 2014). They developed markers of shape variation called response-based imputed predictors to quantify the variation in face shape due to sex and ethnicity, and found that 24 single nucleotide polymorphisms appeared to be associated with differences in these new variables. Limitations of the study are multiple testing of predictor variables and the lack of validation in terms of reproducibility: observers were used to rate the femininity and racial contribution to the facial shape of each individual as a comparison.

Studies using DSM and the principal components distinguished between affected and unaffected individuals using pattern recognition algorithms such as closest mean, linear discriminant analysis, support vector machines and response-based imputed predictors (Claes et al., 2014; Hammond et al., 2005). This involved the generation of random pairs of stratified training-validation subsets of affected-unaffected subjects, with the training set being used to build a dense surface model and the validation set for unseen classification. For example, in closest mean discrimination testing, the average faces of the affected and unaffected individuals in a training set model are generated and the Euclidean dense surface model-based distance between a validation set face and the two means is computed. The validation face is classified as affected or unaffected according to which mean is closest. These algorithms are useful for comparing one homogeneous group with another or with controls. However, they are not useful in detecting facial dysmorphism *per se*. Therefore, a new measurement of face shape was developed and used in the present study (Section 2.7.1).

### 1.3 FACE SHAPE AND EPILEPSY

#### 1.3.1 BRAIN AND FACE DEVELOPMENT

Brain development is closely associated with face development. The central nervous system originates from ectoderm through the process of neurulation in the third week of gestation (Figure 4). Neurulation creates the neural tube, precursor to the brain and spinal cord, and the neural crest cells. By the end of the seventh week of gestation, the cerebral hemispheres have formed. The facial prominences derive from neural crest cells and appear during the fourth week. They include the frontonasal prominence, two paired maxillary prominences and two mandibular prominences. The prominences fuse to form the external face surface by the seventh week. Major facial malformations, such as cleft palate, occur before the seventh week. The process of human face development is described by others in much greater detail (Som and Naidich, 2013).



**Figure 4** An illustration of the origin and development of the human face. Neural crest cells formed after neurulation migrate to form facial prominences which then fuse and develop into the facial soft tissues and skeleton. The figure is reproduced here as permitted by the publisher (Adameyko and Fried, 2016).

Signalling molecules and pathways are complex, with over 80% of all genes being expressed in the mammalian brain (Lein et al., 2007). Despite the complexity, three general principles have been suggested whereby brain development and face development influence each other (Marcucio et al., 2015; Sisodiya, 2008). Firstly, there is a shared origin of face and brain cells from the neural ectoderm. Neural crest cells give rise to the skull, facial skeleton, special sensory organs, and facial soft tissue. They are also subject to the same patterning genes, such as the Homeobox family in the neural tube, as is the developing brain, so that transplanting neural crest cells elsewhere still results in the same structure that it was programmed to form.

Secondly, the brain is a structural scaffold for other tissues developing adjacent to it, including the face. For example, the skull develops in close apposition with the brain, seen in evolutionary and developmental studies. Tensile strain applied by the forebrain cells appears to cause expansion of the developing cranium (Richtsmeier and Flaherty, 2013). In mice, increased brain growth leads to a flatter face whereas a smaller brain is

associated with a longer face and prognathism. This is thought to be from physical displacement of facial tissue sideways by the expanding brain (Marcucio et al., 2015).

Thirdly, there are shared molecular signals between the face and brain. One aspect of face development that is better understood than other regions is the frontonasal ectodermal zone, which controls growth and development of the upper jaw. Blocking a molecule, SHH, from signalling within the brain causes loss of expression of the same gene, *Shh*, in the frontonasal ectodermal zone and thus failure of upper jaw growth in chick embryos (Marcucio et al., 2005). Signalling from the face is also important for brain development and has been recognised more recently. When the rostral neural crest, from which the facial skeleton is made, is surgically excised, there is not only absence of the facial bones, but also anencephaly. Brain development thus seems to be partly dependent on a signal (such as FGF8) from the neural crest (Le Douarin et al., 2012).

Developmental disorders have also suggested a relationship of the face and brain development. Three examples will be described. In craniosynostosis, the primary defect is thought to be premature closure of skull sutures. However, mutations in fibroblast growth factor genes are associated with the condition and these growth factors are known to stimulate neurogenesis and axon growth. In mice models, there appears to be evidence of primary changes in the brain not associated with the effects of suture closure (Richtsmeier and Flaherty, 2013). The first study of genetic polymorphisms and human craniofacial variation found that a fibroblast growth factor gene, *FGFR1*, is associated with the shape of the cranium in normal individuals in three different ethnicities (Coussens and van Daal, 2005). Cranium shape was taken as a simple ratio of width (biparietal diameter) to length (occipitofrontal diameter). The same gene is also implicated in dysembryoplastic neuroepithelial tumours (Rivera et al., 2016).

Another severe developmental disorder is holoprosencephaly. Here the brain and craniofacial structures are both affected by malformations in the midline. In the brain, this ranges from joined cerebral hemispheres with one ventricle to fusion of just the frontal lobe. In the face, cyclopia, the presence of only one eye and orbit, is the most severe malformation but milder anomalies include median cleft lip, flat nose and midface hypoplasia. The most severe brain malformation is strongly associated with the presence of the most severe facial malformation, cyclopia (Ming and Muenke, 1998).

In aniridia, it is well-recognised that there is absence of the iris in affected individuals due to mutations in the *PAX6* gene, a transcription regulator expressed in the developing eye. Only more recently was it recognised that heterozygous *PAX6* mutations may be associated with a range of brain malformations such as polymicrogyria (Sisodiya, 2008) and that *PAX6* is involved in cortical patterning and axon guidance (Georgala et al., 2011).

Much remains to be elucidated but it is clear that the face may be affected in genetic or developmental disorders of the brain.

### 1.3.2 EPILEPSY

Epilepsy is conceptually defined by the International League Against Epilepsy as a “disorder of the brain characterised by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition” (Fisher et al., 2005). A later operational definition states that epilepsy is diagnosed in any one of three conditions: 1. Two or more unprovoked or reflex seizures occurring more than 24 hours apart; 2. One unprovoked or reflex seizure with a high probability of further seizures; 3. Diagnosis of any epilepsy syndrome (Fisher et al., 2014).

Epilepsy may also be classified by aetiology as: Structural, genetic, infectious, metabolic, immune or unknown (Scheffer et al., 2017). Genetic epilepsy includes not just epilepsy due to a known genetic mutation, but also epilepsy thought to have a genetic basis from family history, prior clinical research or molecular genetics.

It is one of the commonest neurological disorders with an annual global incidence of 50.4 per 100,000 (Ngugi et al., 2011). Approximately 75 million people are thought to suffer from epilepsy worldwide, with a prevalence of 11 per 1000 (Ngugi et al., 2010). People with a diagnosis of epilepsy have a two-fold increased risk of premature mortality (Neligan et al., 2011), an increased risk of other psychiatric or somatic disorders (Gaitatzis et al., 2004), and an increased risk of accidents and traumatic injury (Wirrell, 2006). The large psychosocial burden of epilepsy includes decreased intellectual functioning and decreased likelihood of employment or marriage (Sillanpää et al., 1998).

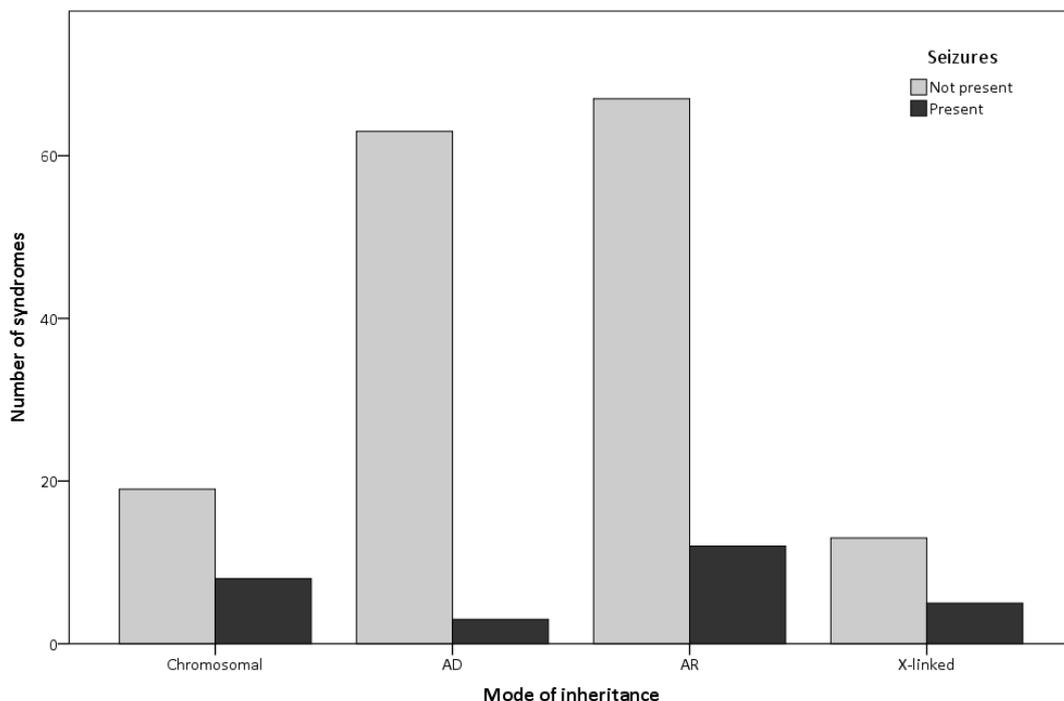
There may be a number of factors affecting face shape in people with epilepsy. These include genetic syndromes, developmental disorders, trauma and anti-epileptic drugs.

### 1.3.3 GENETIC OR DEVELOPMENTAL DISORDERS CAUSING EPILEPSY AND FACIAL DYSMORPHISM

Facial dysmorphism and seizures have been known to co-exist as part of a syndrome for a long time. As outlined earlier, the stigma of epilepsy has been perpetuated by misleading work by Cesare Lombroso amongst others (Section 1.1.3). In the early 1900s, William Aldren Turner, a neurologist at the National Hospital in Queen Square and the then Chalfont Colony for Epileptics, also noted facial similarities. In the population of patients at Chalfont, he thought that 45% had a "facies epileptica". The presence of this facial phenotype was thought to be correlated with the degree of intellectual disability (Eadie, 2006). The patients in institutions were recognised to have more severe epilepsy often associated with intellectual disability or other features

and Turner may have detected facial dysmorphism due to genetic syndromes causing epilepsy. Such syndromes were carefully described once the specialty of clinical genetics developed.

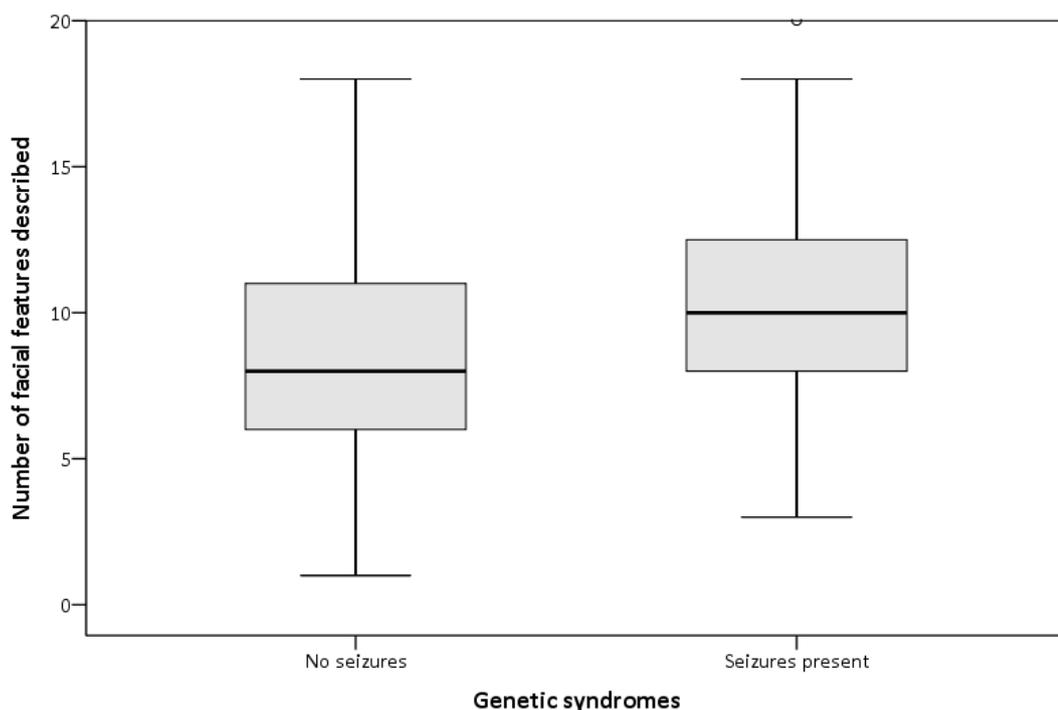
One of the earliest dysmorphology reference books, *Atlas of the Face in Genetic Disorders*, lists all known or suspected genetic disorders with facial dysmorphism, excluding developmental and teratogenic conditions (Goodman, 1977). The book describes 190 syndromes, of which 28 have seizures as a characteristic. These may be subdivided into 8 out of 27 chromosomal disorders (30%), 3 out of 66 conditions with autosomal dominant inheritance (5%), 12 out of 79 conditions with autosomal recessive inheritance (15%) and 5 out of 18 conditions with X-linked inheritance (28%) as summarised in Figure 5. It is notable that chromosomal disorders had the highest frequency of epilepsy. Chromosomal disorders consisted of duplications and deletions of part of or a whole chromosome.



**Figure 5 Genetic syndromes associated with facial dysmorphism according to *Atlas of the Face in Genetic Disorders*, 1977. They are subdivided into mode of inheritance as given in**

***the book and the presence or absence of seizures. AD = autosomal dominant, AR = autosomal recessive.***

Each syndrome is described by a list of all known facial features. This allows for a crude analysis of the number of facial features. In all syndromes associated with seizures, the median number of facial features was 10, whereas it was 8 in the syndromes with no known associated seizures (Figure 6). Chromosomal disorders with seizures were associated with a median of 12 facial features, compared to 7 in autosomal dominant diseases, 8 in autosomal recessive diseases and 9 in X-linked diseases. It is important to note that we now know that the genetic categories are not mutually exclusive and that this may not have been an exhaustive list, but nevertheless it illustrates that (large) chromosomal deletions, duplications or rearrangements may have an association with seizures and facial dysmorphism.



***Figure 6 Genetic syndromes associated with facial dysmorphism according to Atlas of the Face in Genetic Disorders, 1977. The number of described facial features is greater in syndromes with seizures than in those without seizures (10 vs 8 median number of features).***

Thirty years later, in one dysmorphology textbook, 26 syndromes are listed as causing facial dysmorphism and frequent seizures (Jones, 2006). These are shown in Table 2. Again, the majority of the syndromes here are due to structural variants (SVs), which comprise copy number variants (CNVs) as well as chromosomal inversions and translocations (Section 1.3.4).

<b>Syndrome</b>	<b>Cause</b>
Acrocallosal syndrome	Mutation
Angelman syndrome	Structural variant
Autosomal recessive chondrodysplasia punctata	Mutation
Coffin-Siris syndrome	Structural variant*
Deletion 4p syndrome	Structural variant
Deletion 9p syndrome	Structural variant
Deletion 11q syndrome	Structural variant
DiGeorge syndrome	Structural variant
Duplication 3q syndrome	Structural variant
Encephalocraniocutaneous lipomatosis	Unknown
Fetal varicella syndrome	Acquired
FG syndrome	Mutation
Generalised gangliosidosis type 1 syndrome	Mutation
Hypomelanosis of Ito	Structural variant
Killian/Teschler-Nicola syndrome**	Structural variant
Linear sebaceous nevus sequence	Mutation
Menkes syndrome	Mutation
Miller-Dieker syndrome	Structural variant
Neurocutaneous melanosis sequence	Unknown
1p36 deletion syndrome	Structural variant
Schinzel-Giedion syndrome	Mutation
Sturge-Weber syndrome	Unknown
Trisomy 13 syndrome	Structural variant
Tuberous sclerosis	Mutation
X-ATR	Structural variant*
Zellweger syndrome	Mutation

**Table 2** List of syndromes in *Smith's recognizable patterns of human malformation, 6<sup>th</sup> edition, 2006*. Causes are listed afterwards as described in the book. X-ATR = X-linked alpha-thalassaemia/mental retardation. \*Both structural variants and mutations are listed as causes. \*\*Now called Pallister-Killian syndrome.

Looking at the relationship from the field of epilepsy research, some of the same chromosomal syndromes were described in the mid-2000s (Bahi-Buisson et al., 2005;

Singh et al., 2002). A detailed review of 21 chromosomal abnormalities associated with seizures found that ten were associated with microcephaly and five were known to have facial dysmorphism: Angelman syndrome, deletion 2p syndrome, deletion 8q syndrome, Miller-Dieker syndrome, Wolf-Hirschhorn syndrome (deletion 4p syndrome) and fragile X syndrome (Singh et al., 2002). Smith-Magenis syndrome and deletion 9p syndrome were not thought to exhibit facial dysmorphism. It is noteworthy that there is incomplete overlap between syndromes described by neurologists and those by dysmorphologists.

More recent research into epilepsy genetics is described below.

#### 1.3.4 GENETIC EPILEPSY

Epilepsy itself has a genetic basis in at least some individuals, as highlighted by the revised terminology for epilepsy classification (Scheffer et al., 2017). Twin studies show that monozygotic twin concordance rates are 21% for focal epilepsy and 64% for generalised epilepsy, compared to 4% and 9% for dizygotic twins, respectively (Corey et al., 2011). Approximately 5% of people with epilepsy have a first-degree relative and genetic heritability has been estimated to be about 70% for some types (Helbig et al., 2008). The genetic causes may be monogenic or polygenic. Monogenic causes are due to *de novo* or inherited single gene mutations. Polygenic causes may be due to either common variants conferring mildly increased risk of seizures or rarer variants conferring greater increased risk. Some genetic conditions may present as broader syndromes in which seizures are one of many characteristics and these were described in the previous section.

Monogenic epilepsy disorders have been best described in family studies and many of the causative genes are in ion channel subunits. Mutations in such genes may cause different phenotypes in different individuals, with one example being *SCN1A*, the gene encoding sodium channel alpha subunit 1. The clinical phenotype of *SCN1A* mutations

ranges from asymptomatic to generalised epilepsy with febrile seizures plus or to Dravet syndrome (Brunklaus and Zuberi, 2014). This variable expressivity is even true of the same *SCN1A* mutation in the same family and is probably due to the influence of other genes. Another striking feature of mutations in *SCN1A* causing Dravet syndrome is that over 90% occur *de novo* in affected individuals. This can limit the usefulness of a family history and pedigree in making the diagnosis. A high prevalence of *de novo* mutations with reduced penetrance and/or variable expressivity has been documented in other monogenic forms of epilepsy too, both for generalised epilepsy (Pandolfo, 2013) and in focal epilepsy, such as those thought to be due to *KCNT1* gene mutations (Møller et al., 2015).

In most people with a presumed genetic cause of epilepsy, the genes remain unknown. These 'common epilepsies' are those which are not part of a broader syndrome and are thought to show complex, polygenic inheritance (Sisodiya and Mefford, 2011). Six individual genome-wide association studies suggested that there were seven loci for genes associated with the common epilepsies (Leu et al., 2016). A meta-analysis of genome-wide association studies including over 8000 people with epilepsy has also implicated a locus at 4p15.1 (International League Against Epilepsy Consortium on Complex Epilepsies, 2014). Another genome-wide association study in children with febrile seizures either related to or unrelated to the measles-mumps-rubella vaccine also found six loci. Two were known, two were immunological genes and only associated with febrile seizures after vaccination, and two were novel loci for genes involved in magnesium transport and potassium channel regulation (Feenstra et al., 2014). Together, the studies show that common variants may exist but require careful phenotyping and/or large sample sizes for detection.

Rarer variants, with an allele frequency of under 1%, have also been studied. They may take the form of point variants or as structural variants (SVs). Structural variants

include unbalanced chromosomal abnormalities such as deletions and duplications, which are termed 'copy number variants' (CNVs). Structural variants also include balanced abnormalities such as inversions, ring chromosomes or translocations. Oligonucleotide array comparative genomic hybridization (aCGH) is a technique that can rapidly compare a subject's DNA with a reference sample to look for CNVs. With the advent of aCGH, CNVs are the most frequent class of genetic cause of epilepsy. Three particular CNVs – 15q13.3 deletion, 15q11.2 deletion and 16p13.11 deletion – have been associated with genetic generalised epilepsy and the latter two are also associated with focal epilepsy (Sisodiya and Mefford, 2011). All three are also known to be associated with intellectual disability, autism and schizophrenia. Another group of CNVs was found to be causative for epileptic encephalopathies. Overall, CNVs are found in up to 5% of people with epilepsy in some cohorts, such as in children referred for aCGH at one centre (Olson et al., 2014). Copy number variants also demonstrate highly variable expressivity, including within families, such that diagnosis is not straightforward. Even people with epilepsy and structural abnormalities may have SVs contributing to their epilepsy (Helbig et al., 2014; Sisodiya and Mefford, 2011). The finding of seven different CNVs in people thought to have hippocampal sclerosis is a case in point (Catarino et al., 2011). Most other SVs that have been described are not recurrent but instead appear to be rare and such new SVs are continually being found (Mefford et al., 2010). For example, one SV leading to loss of the *WWOX* gene, was found to cause epilepsy and microcephaly (Ben-Salem et al., 2015).

Supporting the theory that shared genes underlie epilepsy and other neuropsychiatric disorders, a genome-wide association study of SVs in genetic generalised epilepsy and controls found that deletions containing genes known to cause neurodevelopmental disorders or autism spectrum disorders were more likely to be found in the epilepsy group with an odds ratio greater than 4:1 (Lal et al., 2015).

Interestingly, further investigation of these SVs reveals additional phenotypic features. One report of 17 people with 15q13.3 deletion found that there was no significant dysmorphism, but that six (35%) had a bulbous nasal tip (van Bon et al., 2009). A subsequent review of 260 cases with 15q13.3 deletion again found no consistent dysmorphic features, but found that 10 of 55 children were reported to have microcephaly and 15 of 55 were reported to have macrocephaly (Lowther et al., 2015). In a European study, rare CNVs have been found in 71 out of 222 (32%) of children with 'complex' epilepsy associated with dysmorphism, brain abnormalities on MRI or intellectual disability (Helbig et al., 2014). The authors had clinical information on the presence or absence of dysmorphism for 207 subjects and the presence of dysmorphism was associated with a significantly increased chance of having a rare CNV (50% in those with dysmorphism vs 21% in those without). Unfortunately, the nature of dysmorphism is not clarified.

Whole exome and whole genome sequencing, also called next generation or massively parallel sequencing, is a technology which automates DNA sequencing using large-scale high-throughput machines. It has been used successfully to identify new genetic syndromes with epilepsy. For example, *EEF1A2* mutations have been found to be the cause of intellectual disability, epilepsy, autism and microcephaly in four people (Nakajima et al., 2015). Two of the four, both unrelated girls, had facial dysmorphism with deep set eyes, epicanthal folds, depressed nasal bridge, lower lip eversion and downturned corners of the mouth. A limited form of next generation sequencing can be used to explore only the commonest genes involved in a condition. Such targeted sequencing of 265 genes has been successfully used in a gene panel for epilepsy with a detection rate of nearly 50% (16) in 33 selected patients from across Germany and Switzerland (Lemke et al., 2012). It is not clear if facial dysmorphism was present but the authors note that clinical assessment of dysmorphism is not helpful except in specific cases. In the UK, gene panel testing for epilepsy syndromes is becoming

available (Dixit and Suri, 2016). Note that currently aCGH is still necessary for confirming the presence of SVs.

To my knowledge, there is only one clinical overview of facial dysmorphism in epilepsy (Dixit and Suri, 2016). This educational article reviews nine recently recognised syndromes, specifically chosen because they are not detected by karyotyping. Three of the syndromes are due to structural variants and found using aCGH: 1p36 deletion, Koolen-deVries syndrome and Kleefstra syndrome.

### 1.3.5 BENEFITS OF GENETIC DIAGNOSIS

Given ongoing research, the cost and the delay involved in reaching a genetic diagnosis in epilepsy, it is pertinent to ask whether it is necessary to obtain such a diagnosis. In the case of CNVs in people with hippocampal sclerosis, for example, surgical resection was successful in all eight described (Catarino et al., 2011). It remains true that the complete phenotypes and significance of many structural variants are not fully known.

Nevertheless, there are important reasons for pursuing a genetic diagnosis. Firstly and most importantly, it may improve clinical management. It may end the 'diagnostic odyssey' which in itself may be therapeutic for patients and may prevent unnecessary investigations. A diagnosis may also help for medicolegal and employment purposes, as well as allowing patients to seek further information or support. Lastly, a genetic diagnosis often aids prognosis and prenatal counselling (or preimplantation genetic diagnosis). In a consecutive series of 88 people attending an adult genetics clinic in the USA, 52 were referred to appropriate specialties for related diseases and 35 benefited from symptom management (Maves et al., 2007). Out of 53 patients with a provisional diagnosis, 27 had it changed in clinic with 11 changed from the history and physical examination alone.

In epilepsy, a genetic diagnosis can optimise treatment such as avoidance of carbamazepine and lamotrigine in *SCN1A* mutation-positive Dravet syndrome or the use of the ketogenic diet in *GLUT1* deficiency syndrome (Brunklaus et al., 2012; Kass et al., 2016). A case series by Kasperavičiūtė and colleagues describes a patient with seizures, who was found to have 17q12 deletion (Kasperavičiūtė et al., 2011). The diagnosis allowed him to receive specific specialist care regarding known consequences of the deletion which included impaired glucose tolerance, renal cysts and hypomagnesemia. A genetic diagnosis of epilepsy in itself is beneficial to patients and families. In a written survey sent to parents by Caminiti and colleagues, 71% of parents still wished to have a genetic diagnosis for their affected child even when told it would have no effect on the child's medical care (Caminiti et al., 2016).

### 1.3.6 TRAUMA

Trauma may affect the face in people with epilepsy and thereby affect analysis using face shape. Individuals with epilepsy have a 1.6 times greater risk of accident (absolute risk of 17% per year) than matched controls without epilepsy according to a prospective European study in which participants kept a daily diary of accidents (van den Broek and Beghi, 2004). There were significantly more wounds, abrasions and concussions in people with epilepsy and these were mostly related to seizure activity. It is not recorded how many of these affected the face but others report that about 80% of injuries are to cranial soft tissues (Lawn et al., 2004). Other reviews suggest that 5-10% of people with epilepsy may sustain dental or jaw injuries each year (Wirrell, 2006). The type and frequency of seizures are known to affect the frequency of injuries, with drop attacks and generalised tonic-clonic seizures being the most likely to lead to injury (Lawn et al., 2004; Tiamkao et al., 2009). Studies were often conducted in selected non-representative groups of people with epilepsy, such as those presenting to tertiary centres, those in nursing homes or those with intractable seizures. Previous facial morphometric studies have ignored facial injuries or excluded such cases on the

basis of patients' recall of injuries (Evison et al., 2010; Hammond et al., 2005; Kau et al., 2010).

### 1.3.7 ANTI-EPILEPTIC DRUGS

The so-called first generation anti-epileptic drugs are known to cause adverse effects that can alter face shape. These drugs include carbamazepine, phenobarbital, phenytoin and sodium valproate (Perucca, 2005). Facial dysmorphism may be especially severe for a fetus exposed to maternal phenytoin use with midface hypoplasia and other features in what some have called the "anticonvulsant face" (Orup et al., 2003). This may be of relevance given that some people with epilepsy may have suffered from *in utero* exposure to phenytoin, for example in familial epilepsy syndromes.

In children and adults using first-generation anti-epileptic drugs, more subtle differences may occur. An analysis of patients with epilepsy followed up for 11 months showed that 50 out of 355 patients (14%) had skin and mucosal changes (Collaborative Group for Epidemiology of Epilepsy, 1988). Changes included gingival hyperplasia, acne and hirsutism. The study also showed that 136 patients (38%) experienced somnolence, vertigo, unsteadiness, ataxia or diplopia. These adverse effects may of course further increase the risk of facial injury in people with epilepsy, and are also noted in many newer anti-epileptic drugs.

Phenytoin has been described to cause coarsening of facial features separate to the skin changes noted above. A case report of two twin pairs, with one of each pair using long-term phenytoin, noted thickening of the lips and nose (Falconer and Davidson, 1973). Facial coarsening occurred in the majority of women using it according to one Brazilian report, but details are inconsistent (Trevisol-Bittencourt et al., 1999). There is limited information on the characteristics or, indeed, definition of such coarsening.

A more recent single centre study of cosmetic side effects due to anti-epileptic drugs in 1900 patients found that 2.5% of people taking phenytoin experienced gingival hyperplasia but other skin changes were not significantly increased (Chen et al., 2015). Sodium valproate was associated with significantly increased risk of hair loss. Both sodium valproate and pregabalin were associated with weight gain. The authors did not find any significant cosmetic effects from newer anti-epileptic drugs but note that larger scale studies are necessary.

Weight gain indirectly affects face shape and has been well recognised in the case of sodium valproate. About 50% of people taking the drug show significant weight gain (>5kg) with risk factors being female sex and duration of therapy (Verrotti et al., 2011). Similarly, it would be expected that drugs known to cause weight loss, such as topiramate, may also affect face shape.

Therefore, anti-epileptic drug use may affect face shape by different means and they should be considered as potential confounders in the study of face shape in people with epilepsy.

## 1.4 AIMS

In summary, face and brain development are closely linked and this relationship is partly mediated by genetic factors. Therefore, disorders of brain development or function, such as epilepsy, may affect face shape too via a shared genetic mechanism. However, the clinical study of face shape is still primarily subjective and underused by the majority of physicians, including neurologists. There are powerful methods available to capture face shape information, such as stereophotogrammetry and sophisticated tools for analysis of these data, such as the use of dense surface modelling and principal component analysis. These methods may be useful in objectively detecting changes in face shape in people with epilepsy caused by particular genetic or developmental abnormalities.

### 1.4.1 VALIDITY OF STEREPHOTOGRAMMETRY AND DENSE SURFACE MODELLING

I hypothesise that stereophotogrammetry and dense surface modelling are reproducible and robust techniques for face shape analysis in a population of subjects with epilepsy.

#### Aims

- To assess intra-operator and inter-operator reproducibility with these techniques.
- To compare dense surface modelling to landmark-based analysis of facial shape and subjective assessments of dysmorphism.
- To determine the effect of camera calibration, camera orientation and facial expression on the resulting image quality.
- To explore the effect of body mass index on the analysis of face shape and attempt to correct for this.

### 1.4.2 INVESTIGATION OF FACE SHAPE IN PEOPLE WITH EPILEPSY

I hypothesise that face and brain development are closely co-ordinated such that structural variants and developmental brain lesions that are associated with epilepsy are also associated with changes in facial shape or symmetry.

#### Aims

- To assess if facial shape change is related to the presence of established pathogenic structural variants and explore the function of genes contained in such variants.
- To explore how facial shape change may be related to the presence of, the type of, or the size of a structural variant.
- To determine if dense surface models may be useful in determining atypical face shape in people with a particular genetic syndrome.
- To explore the relationship of facial symmetry and shape to the presence of developmental brain lesions.

## 1.5 EXPERIMENTAL APPROACH

The first part of the study is to validate the techniques of stereophotogrammetry and dense surface modelling for face shape analysis. To do this, I propose to analyse each step in the process. Firstly, using a group of controls, I will explore reproducibility for identification of landmarks on the face, traditional measurements of face shape, and finally for two new metrics of face shape change. Then I will assess how robust the techniques are to environmental variables including head position, facial expression and camera re-calibration.

People with epilepsy or specified genetic syndromes will then be recruited for the study over a period of 2-3 years. This will include adults with capacity to consent, children with the consent of parents or guardians, and adults without capacity to consent for whom the assent of a consultee has been obtained. Many of these people will already have been recruited as part of a related study of the genetics of epilepsy. Participants will undergo face image capture at an epilepsy treatment centre, height and weight measurement, and a blood test for genetic testing, if not already performed. Clinical details will include age, sex, handedness, history of facial trauma, previous anti-epileptic drug use, genetic diagnoses, imaging findings and neuropsychometry findings. These will be recorded at the time of recruitment or from accessing the medical record. Some participants may be recruited for image capture only a second or third time to assess for the effects of age.

After recruitment, I will perform dense surface modelling of the 3D face images to analyse face shapes using two new metrics – one for overall shape relative to controls and one for asymmetry. As they are new, I will explore the relationship of these metrics to the physiological variables of age and body mass index. I will also assess how well these metrics can differentiate between participants with underlying genetic disorders, as classified by the presence of structural variants, and participants without such

variants. I will then assess if the metrics can differentiate those people with epilepsy with a developmental brain lesion and those people with epilepsy without such a lesion. Finally, I hope to recruit a group of people with a known genetic syndrome to see if stereophotogrammetry and dense surface modelling can be used to detect shared facial features.

## 1.6 JUSTIFICATION OF RESEARCH

The experimental approach described above fulfils the aims of this thesis. I believe that it is important to investigate face shape change in people with epilepsy. As I have described, face shape assessment is already commonly used in medicine, especially clinical genetics, to help diagnose genetic disorders. In some people with epilepsy, genetic causes are well-recognised and novel structural variants and genetic mutations have been recently discovered. Face shape analysis is part of the detailed phenotyping that will assist in understanding the effect of such conditions. The new method I propose involves digital image capture in two milliseconds and an objective, quantitative measurement using detailed 3D face shape images, which will be matched for age, sex and ethnicity to a reference cohort. The technique I investigate in my study is arguably less unpleasant and/or more objective than current techniques including direct anthropometry and inspection by a clinician.

As noted above and in Section 3.1, there are few detailed studies of reproducibility and accuracy of stereophotogrammetry and landmark placements. There are no known studies of reproducibility and accuracy of dense surface modelling. Therefore, it is important for me to validate the technique I propose to use before recruitment of subjects. I will do this validation study (Chapter 3) only for adults with the capacity to consent. Some of these adults have epilepsy, which is important as they are the target population of my study.

For all later studies, it is important to include participants who lack capacity to consent. I have described earlier (Section 1.3.4) that the presence of structural variants is associated with intellectual disability. I will not be able to assess the effect of a range of structural variants on facial shape thoroughly if participants who lack capacity to consent are not included. The same may apply to the presence of developmental brain

lesions, many of which are associated with developmental delay and intellectual disability (Guerrini and Dobyms, 2014).

Similarly, many of those who are diagnosed to have genetic disorders are children and the study would be limited in power by excluding children. However, all children had already had genetic testing as part of their clinical care or other research studies and they will only undergo image capture as part of this study. Image capture will only be performed after informed consent from the parents in their local language and in accordance with local ethics committee or institutional review board approval.

For all participants, after orientating and uncovering the face, image capture will use a single camera flash. The only invasive procedure in a minority of subjects will be a blood test for genetic analysis with their consent. Children will only be recruited if they have already had such a test performed for clinical reasons. In adults without capacity to consent, such a blood test was performed at the same time as their clinic visit and clinically indicated blood sampling.

Face images and collected clinical or genetic data are kept on an encrypted drive in a locked room at all times. Face images are not used for any other purpose except for the following study. Storage of images is necessary for later analysis with other subjects and verification of findings.

## CHAPTER 2: METHODS

### 2.1 LITERATURE REVIEW

A literature review was conducted for each chapter. The principal search database used was MEDLINE via Pubmed, developed by the National Institutes of Health. The OVID database and was also consulted when appropriate, and occasionally, older articles were retrieved from the JSTOR collection.

Multiple search terms and phrases were used for each chapter and section. As one example, for Section 1.1.4, the following terms were used:

- Clinical genetics
- Craniofacial
- Dysmorphic
- Dysmorphism
- Dysmorphology
- Facial gestalt
- Malformations
- Teratology

### 2.2 RECRUITMENT OF SUBJECTS

#### 2.2.1 REFERENCE SUBJECTS

Reference control subject images were used to provide a database of age-matched, sex-matched, ethnicity-matched images for many of the subsequent analyses. Reference subjects were part of most study groups, but they were not compared to any study group except for Group 7. Individual reference subject images were only used in a study of inter-operator reproducibility (Sections 3.3.3 and 3.4.2) and not analysed further. However, composite average reference faces were created in a process

described later (Section 2.7.1) and these average images were used to assess facial shape change. In group 7, composite average reference faces were also compared to each of the probands (Chapter 8). All reference subject images were provided for use in the DSMs by Peter Hammond at the UCL Institute of Child Health, London.

Reference subjects were all white European adults and children. They were previously recruited as volunteers, unaffected relatives of patients, or healthy infants attending a routine postnatal clinic, from the UCL Institute of Child Health. All reference subjects were healthy and had no known syndrome, previous craniofacial surgery or trauma. They have been previously used for comparison to children and adults with Williams syndrome, Smith-Magenis syndrome, 22q11 deletion syndrome, Noonan syndrome, Fabry disease, Wolf-Hirschhorn syndrome, fibrodysplasia ossificans progressiva and fetal alcohol syndrome (Cox-Brinkman et al., 2007; Hammond, 2007; Hammond et al., 2012a, 2012b, 2005, 2004; Suttie et al., 2013).

For reference subjects, images were captured using the DSP400 and MU2 commercial cameras (3dMD; Atlanta, GA, USA). There is no significant difference between face images captured on different stereophotogrammetric cameras (Weinberg et al., 2006). Multiple cameras have been used in previous studies (Hammond et al., 2005).

## 2.2.2 RECRUITMENT OF CONTROL SUBJECTS

Twenty control subjects were recruited to assess reproducibility of stereophotogrammetry and dense surface modelling under different conditions, including face position relative to the camera, face expression and camera calibration (Section 3.3.2). They were all adult volunteers, who were able to co-operate for repeated image acquisition. Eight of them were also people with epilepsy who had been recruited for the main study. Note that none of these 20 controls were compared to any other group and were used solely to assess reproducibility. No control subject had previous significant facial injury or surgery.

### 2.2.3 RECRUITMENT OF PEOPLE WITH EPILEPSY IN THE UNITED KINGDOM

Principal recruitment of subjects was from routine attendance at adult epilepsy clinics in two locations within University College London Hospitals NHS Foundation Trust: the National Hospital for Neurology and Neurosurgery in Queen Square, London and at the Chalfont Centre for Epilepsy in Chalfont St Peter, Buckinghamshire. Subjects were also recruited from inpatient wards in these two locations, named the Sir Jules Thorn EEG Telemetry Unit and the Sir William Gowers Epilepsy Assessment Unit, respectively. The inclusion criteria consisted of participants who were:

- Over the age of 18 years
- Undergoing investigations for epilepsy, or have a known diagnosis of epilepsy
- Able to provide informed consent
- Willing to participate

Recruitment of people with *HNF1B* mutations occurred at another site, the UCL Institute of Child Health.

### 2.2.4 RECRUITMENT OF PEOPLE WITH EPILEPSY IN EUROPE

Three other sites of recruitment for children and adults with epilepsy were also used:

- Meyer Children's Hospital (Ospedale Pediatrico Meyer), Florence, Italy
- Erasmus Hospital (Hôpital Erasme), Brussels, Belgium
- University Hospitals Leuven, Campus Gasthuisberg (Universitaire ziekenhuizen Leuven), Leuven, Belgium

For each site, the study had been approved by the relevant ethics committees and/or institutional review boards. Written informed consent was obtained from study participants or informed assent was obtained from parents according to national

standards. Participants were either patients at the above institutions, or relatives of patients.

### 2.2.5 GROUPS USED FOR ANALYSIS

In total, 869 subjects were recruited for stereophotogrammetry and dense surface modelling. These subjects included control subjects, people with epilepsy and relatives of people with epilepsy. People with epilepsy comprised those with and without genotype data, those with and without brain imaging data, those with and without body mass index data, and those with and without age or ethnicity matched control subjects. Therefore, they were incorporated into a number of different groups, depending on the individual study. In some studies, facial shape change was quantified using a novel measurement called ‘Face Shape Difference’ (FSD). FSD required all recruited subjects to be compared to matched reference faces. Therefore, reference subjects were included in groups using FSD in subsequent analysis. If there were insufficient reference faces matched for age or ethnicity for a particular subject, then that subject was excluded. The study groups are described below and a summary is shown in Table 3.

Groups	1	2	3	4	5	6	7
Centres	All		UK only				
Relatives	No		Yes	No	Yes	No	Yes
Children	Yes	No					Yes
Genotyped	Yes	N/A					
BMI data	N/A	Yes			N/A		
MRI data	N/A					Yes	N/A
FSD used	Yes	No	Yes			No	Yes
rFSD used	No					Yes	

**Table 3 Summary of characteristics of the different groups used in studies. The purpose of this groups is described in the main text. ‘Centres’ refers to the sites of recruitment and was either from the UK only (‘UK’) or from the UK and other European sites (‘All’). ‘Relatives’ is the inclusion of controls and unaffected relatives in analysis – they are separate from the reference subjects who were not used in any analysis but were part of all studies using a quantitative measure called ‘FSD’ (Section 2.7.1). ‘Children’ is the inclusion of people under 16 years of age for analysis. ‘Genotyped’ means only people with genotype data were included in the group. ‘BMI data’ and ‘MRI data’ means that only people with BMI or MRI**

*imaging data, respectively, were included in the group. 'FSD used' refers to analysis of face shape using FSD within that group. 'rFSD used' refers to the use of reflected face images and the reflected FSD for all subjects to assess face asymmetry.*

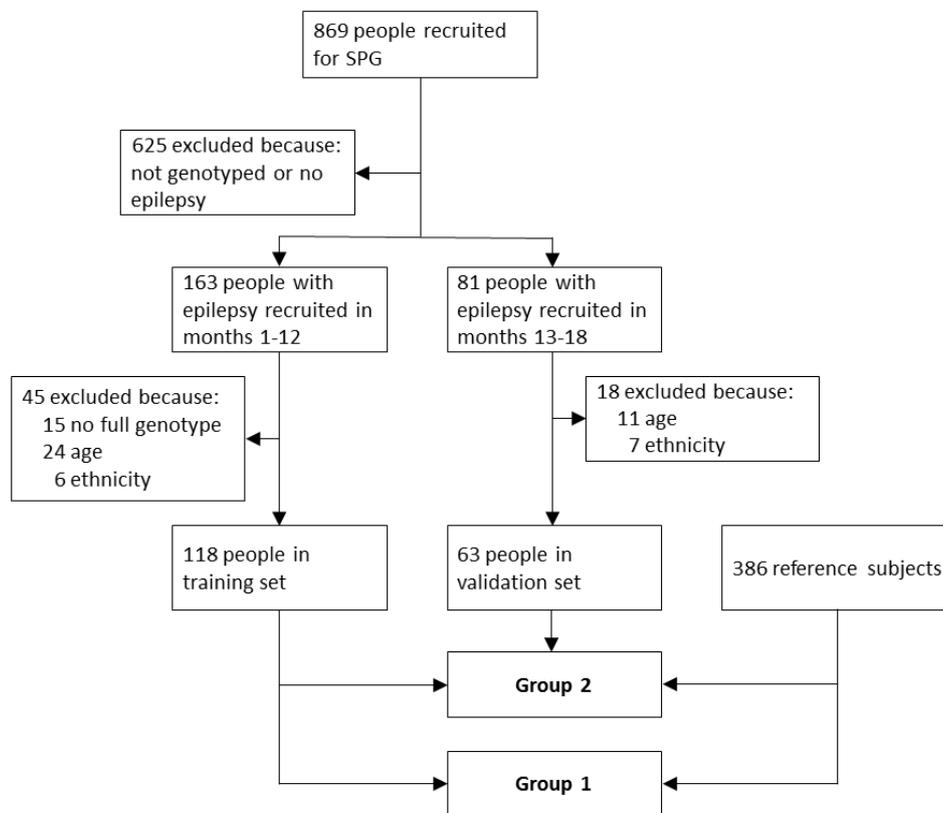
Group 1 (Figure 7): This group was used as part of validation studies (Chapter 3). It was then used to study the relationship of facial shape change to the presence of pathogenic structural variants (Chapter 4). A positive result was seen and the group was then also incorporated as a training set within group 2, to see if the result could be validated in an additional sample.

Inclusion criteria:

1. Subjects who had undergone genotype analysis from two UK centres (Section 2.2.3) and three European centres (Section 2.2.4) over one year of recruitment.
2. Children with suspected epilepsy or a formal diagnosis of epilepsy made by a paediatric neurologist. Adults with a diagnosis of epilepsy made by a trained neurologist specialising in epilepsy.

Exclusion criteria:

1. People with known Mendelian inheritance of epilepsy or a syndrome associated with epilepsy.
2. People with known imbalances of a whole chromosome were excluded.
3. People who could not be compared to reference subjects for quantification of facial shape change (i.e. FSD) due to lack of matched controls. This could be due to age (insufficient matched controls existed over 52 years for men and over 54 years for women) or ethnicity (sufficient matched controls only existed for white Europeans).
4. Further exclusions were made *ad hoc* for sensitivity analyses, described in Chapter 4.



**Figure 7 Study recruitment and groups used in exploring the relationship of face shape to structural variants in people with epilepsy. Initially, 118 people were analysed in group 1. An additional 63 people were then analysed separately within group 2 for validation. Reference subjects were only used to calculate FSD. SPG = Stereophotogrammetry**

Group 2 (Figure 7): This group comprised all subjects in group 1 and additional subjects recruited using the same inclusion and exclusion criteria over a further six months. The group 1 subjects formed a training set and the additional subjects formed a validation set. The results found using group 1 earlier were validated using only the new, additional subjects. Group 1 had to be included in group 2 to enable validation (described in Section 4.4.3), but the training set and validation set did not overlap. The whole of group 2 was then also used to assess the effect of three aspects of the size of pathogenic SVs on face shape: size of the SV in base pairs, size of the SV in number of genes, and size of the SV in number of genes expressed in the developing human

forebrain (Chapter 5). Within Chapter 5, a subsequent analysis was performed of only those people with complete information on all detected SVs, which is termed Group 2B (described separately in Chapter 5).

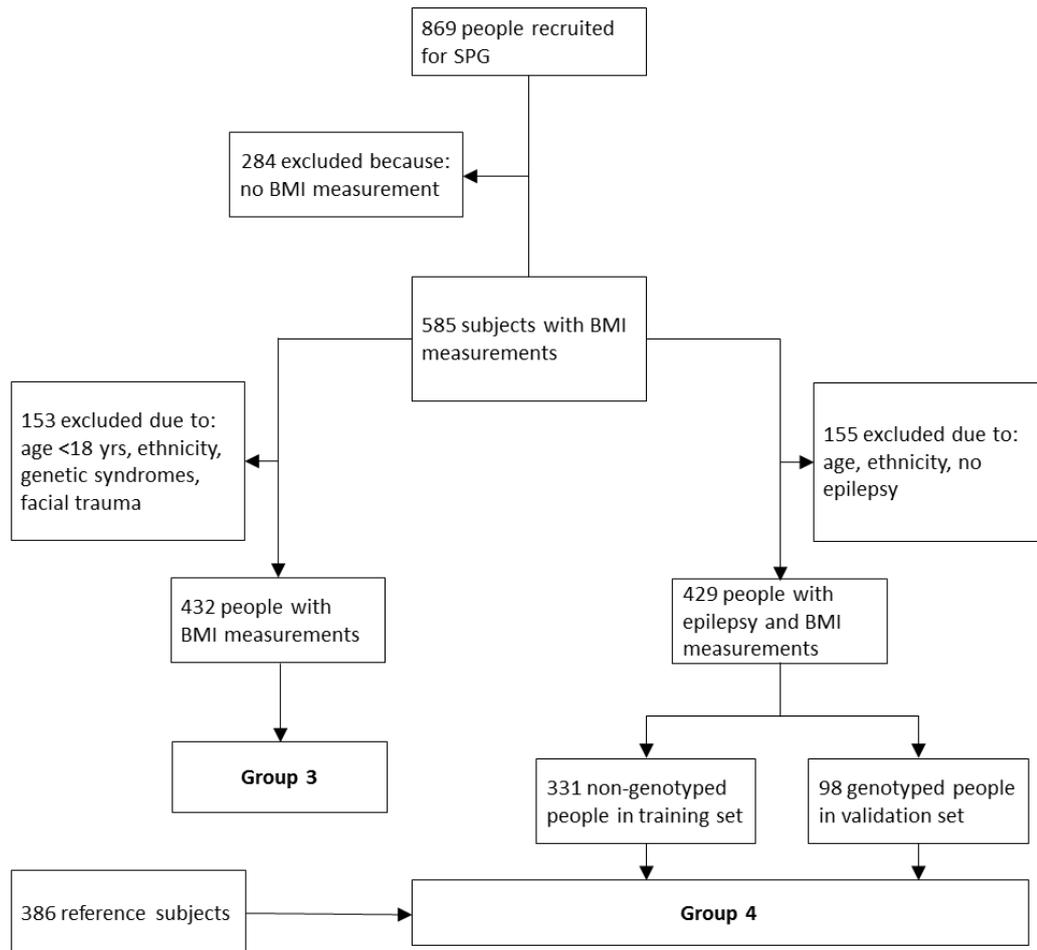
Group 3 (Figure 8): This group was used in further validation studies involving facial expression, camera angle and camera calibration (Chapter 3). It was then used to explore the effect of body mass index (BMI) on face shape (Chapter 6).

Inclusion criteria:

1. People with height and weight measured at the time of stereophotogrammetry.

Exclusion criteria:

1. People under 18 years old
2. People who were not white Europeans. There were insufficient subjects to explore how face shape changes with BMI in other ethnicities.
3. People with craniofacial deformity, with previous craniofacial surgery, or with any known genetic syndrome/diagnosis were excluded.



**Figure 8 Study recruitment and groups used in exploring the change in face shape to body mass index in different subjects. The first study used group 3 and found a predictable relationship between face shape and body mass index. The second study used group 4 to see if a quantitative measurement of face shape change, FSD, could be adjusted for BMI and if doing such an adjustment would affect the accuracy of previous findings in people with epilepsy and structural variants. SPG = Stereophotogrammetry**

Group 4 (Figure 8): The study using group 3 showed a predictable change in face shape with change in BMI. This new group was used to adjust the measure of face shape change (FSD) to account for BMI (Section 6.4.6). In particular, FSD was adjusted in people who had been genotyped to see if previous results in Chapters 4 and 5 were still true without BMI as a confounder.

Inclusion criteria:

1. People with height and weight measured at the time of stereophotogrammetry.

2. People with a diagnosis of epilepsy made by a trained neurologist specialising in epilepsy.

Exclusion criteria:

1. People who could not be compared to reference subjects for quantification of facial shape change (i.e. FSD) due to lack of matched controls. This could be due to age (insufficient matched controls existed over 52 years for men and over 54 years for women) or ethnicity (sufficient matched controls only existed for white Europeans).

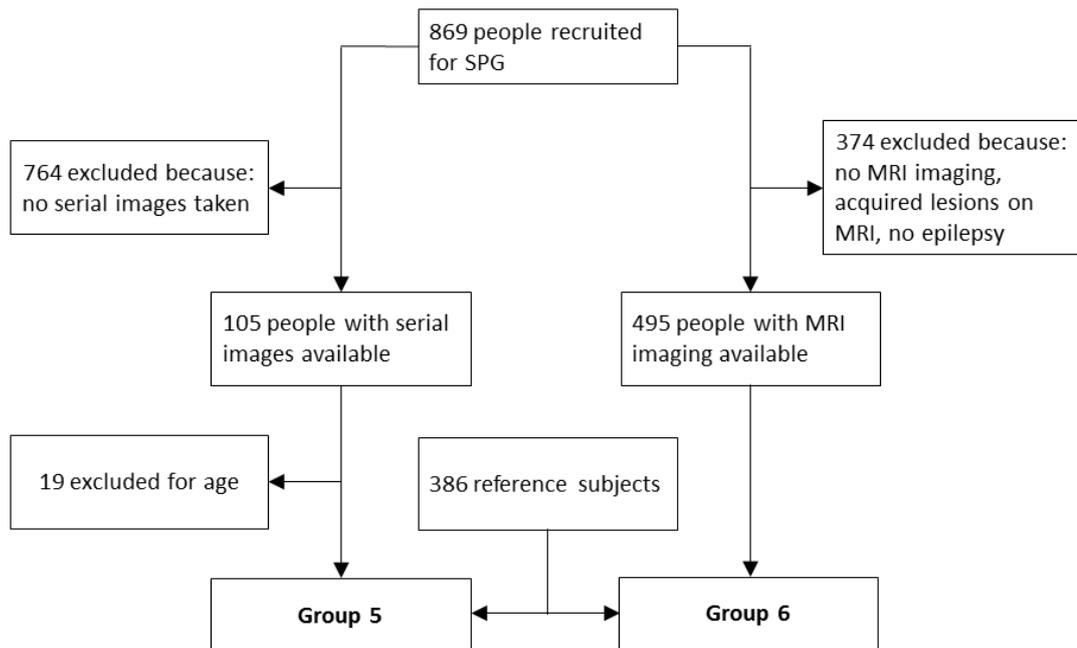
Group 5 (Figure 9): This group was used in a longitudinal study explored the effect of age on face shape during the study recruitment period (Section 6.4.7), by assessing serial images of the same subject.

Inclusion criteria:

1. People with a diagnosis of epilepsy made by a trained neurologist specialising in epilepsy.
2. People with more than one image taken in separate stereophotogrammetry sessions at least two weeks apart.

Exclusion criteria:

1. People who could not be compared to reference subjects for quantification of facial shape change (i.e. FSD) due to lack of matched controls. This could be due to age (insufficient matched controls existed over 52 years for men and over 54 years for women) or ethnicity (sufficient matched controls only existed for white Europeans).



**Figure 9 Study recruitment and groups used in two studies. The first is a longitudinal assessment of face shape change, using FSD with time, in group 5. The second study uses group 6 and assesses face shape asymmetry in people with epilepsy and brain lesions compared to people with epilepsy and without any brain lesions. SPG = Stereophotogrammetry**

Group 6 (Figure 9): This group was used in a study of the relationship between facial symmetry and the presence of unilateral or bilateral developmental brain lesions.

Inclusion criteria:

1. People with a diagnosis of epilepsy made by a trained neurologist specialising in epilepsy.
2. People who had undergone magnetic resonance imaging of the brain for investigation of their epilepsy at the UK sites and had these scans reviewed by a neuroradiologist.

Exclusion criteria:

1. People with MRI imaging abnormalities including acquired intracranial lesions (such as stroke, trauma), undefined lesions or multiple lesions.

2. Exclusions were made *ad hoc* for sensitivity analyses, described in Chapter 7.

Group 7: This group was used in a study of subjects with one particular genetic condition. It comprised only subjects with *HNF1B* mutations, including deletions, who were recruited during a support group event at the UCL Institute of Child Health.

Inclusion criteria:

1. People with a confirmed genetic diagnosis of *HNF1B* mutation or deletion.

Exclusion criteria:

1. People who could not be compared to reference subjects for quantification of facial shape change (i.e. FSD) due to lack of matched subjects. This could be due to age (insufficient matched reference subjects existed over 52 years for men and over 54 years for women) or ethnicity (sufficient matched reference subjects only existed for white Europeans).

## 2.3 IMAGE CAPTURE

Image capture was performed using the Mirror Vectra CR10 software program (Canfield Scientific, Fairfield, NJ, USA). The hardware used in all cases was a standard laptop or desktop personal computer, attached to the Canfield Vectra CR10 3D Face (Canfield Scientific, Fairfield, NJ, USA) camera system, shown in Figure 10. The manufacturer's specification sheet states that the camera has a resolution of 10 megapixels and resolves surfaces to 1.1mm. Capture time is two milliseconds. The camera was calibrated. Calibration involves placing a rigid patterned board in front of the camera at two angles and capturing two images of it. The camera uses this information to optimise creation of 3D images. Calibration was conducted at the start of each image capture session and after any movement of the camera, as per manufacturer's recommendations. The camera lenses were not in the horizontal plane

and were angled upwards by 20 degrees (i.e. the camera system was rotated 20 degrees anticlockwise around the long axis of the system). This has been well-documented in other stereophotogrammetry studies and aids capture of the inferior aspect of the nose and chin (Ghoddousi et al., 2007; Heike et al., 2010). The angle is illustrated in Figure 11.

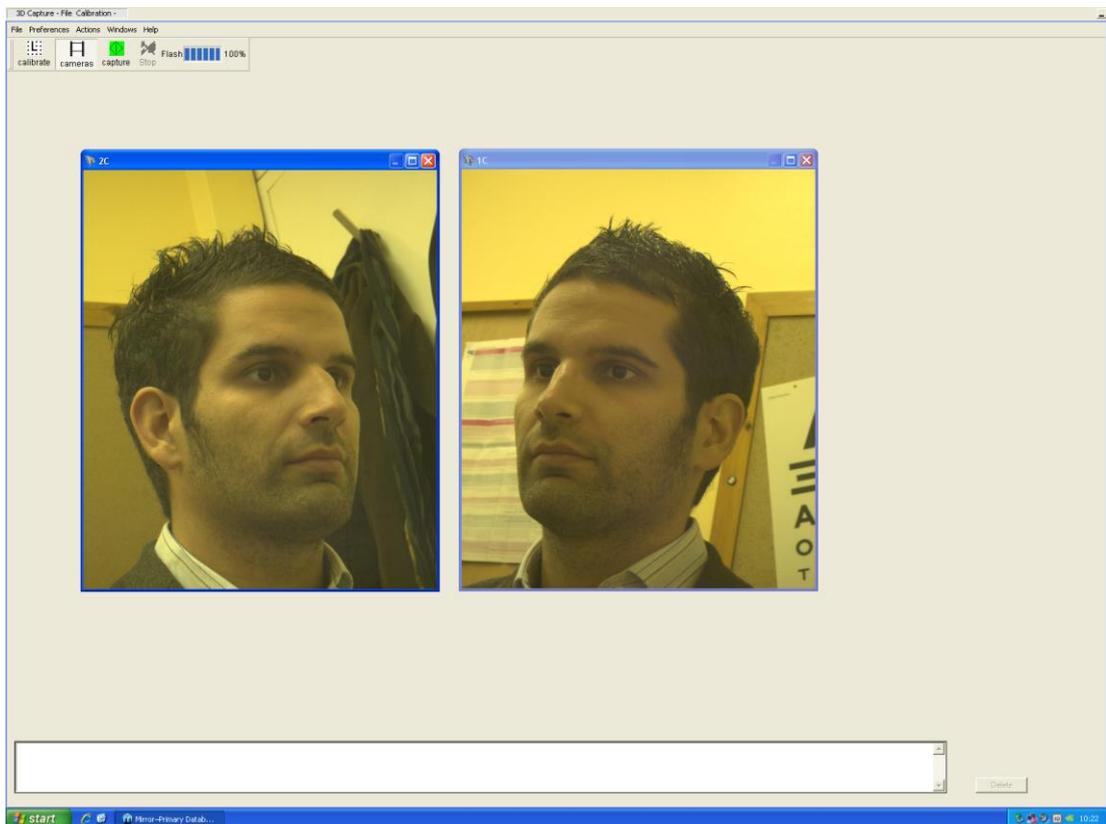


***Figure 10 The Canfield Vectra CR10 stereophotogrammetry camera system. The image shows two flash units above the main camera body. The camera body has two pods on each side with camera lenses facing along the axis of the camera. A mirror is used to allow the lenses in each pod to capture subjects in front of the camera. This arrangement allows the camera to be relatively flat and hence portable. Image is used with the written permission of Canfield Scientific.***

All participants were seated in a room with adequate indoor lighting. A height-adjustable swivel chair was provided and participants sat facing the camera approximately one metre away. A typical position for image capture is shown in Figure 11. Participants' faces were then aligned within the two on-screen target areas (Figure 12). If it was not possible to do this by adjustment of the chair, then the camera tripod height was adjusted too. Visual inspection was used to minimise lateral rotation of the face by rotating the swivel chair.



**Figure 11** A typical position used for image capture of a subject's face from another study using the same Canfield Vectra CR10 3D stereophotogrammetry camera model. Image is reprinted from Ghoddousi et al., 2007 with permission from Elsevier.



**Figure 12** Screenshot from the Mirror Vectra CR10 software program illustrating how subjects are aligned within the target field for each of the pairs of cameras to ensure

*adequate coverage of the face surface in the 3D images (permission obtained from subject for use).*

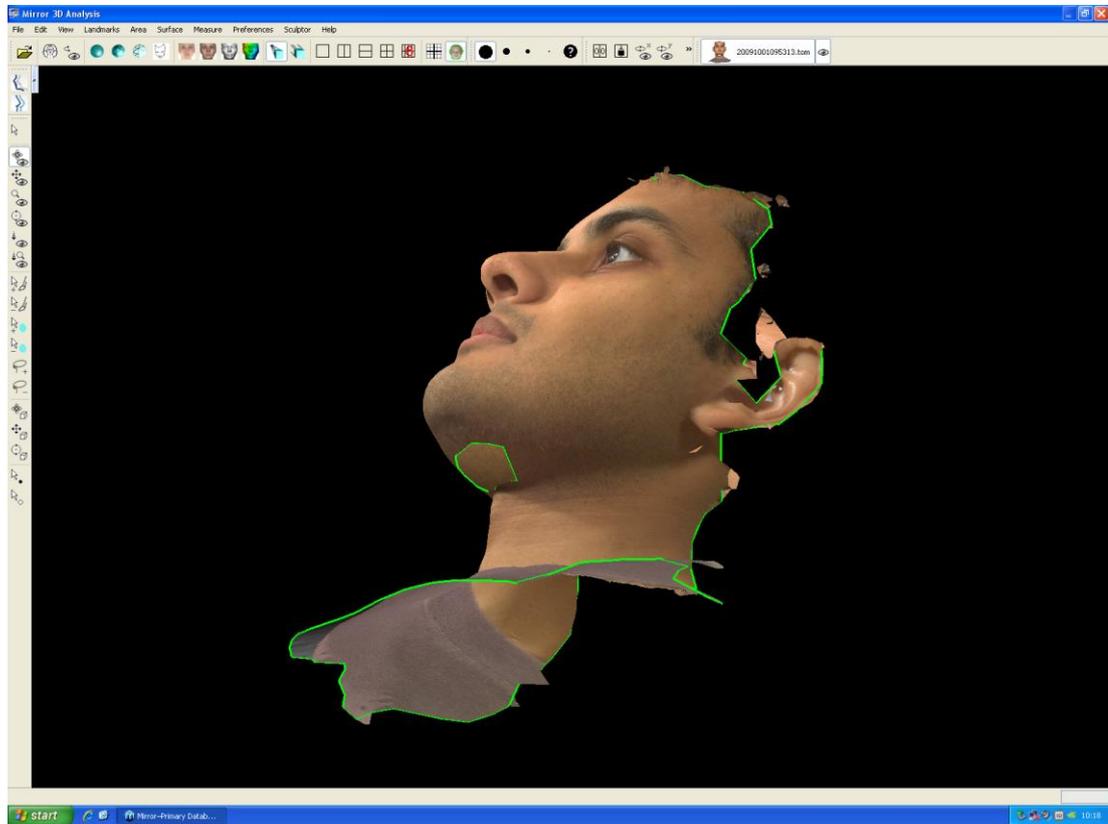
Participants were then asked to remove any coverings on their head or garments that obstructed their face, including the chin and upper neck. Earrings and other piercings were removed if practical. They were also asked to adjust their hair to uncover their forehead and ears as much as possible. Hair bands and clips were provided for this purpose.

All subjects were told to look at a target straight ahead in the vertical midline and slightly above their eye level. Height was adjusted so that the top of the camera was at eye level. Participants were told not to smile and to keep their lips together and their eyes open. If this was not possible, for example, in young children, multiple images were taken and the closest approximation to the neutral expression was used. An operator checked that all parts of the face were visible to the camera pairs using the software program (Figure 12) and then the image was taken. The procedure is in accordance with previous recommendations by one team from their experience of clinical stereophotogrammetry (Heike et al., 2010).

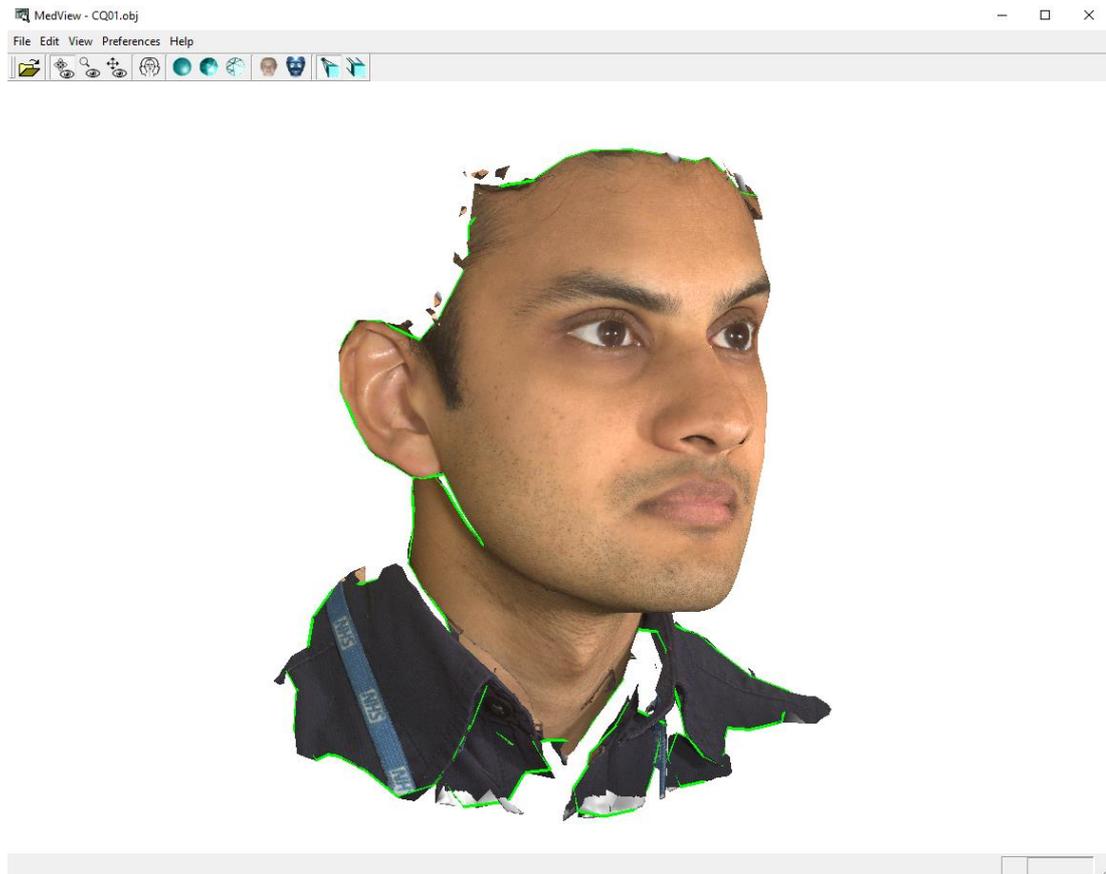
Image capture is triggered from the personal computer and uses a synchronous flash from two on-board flash units. Images were assessed for eye opening, expression and adequate coverage of the face surface. Three-dimensional (3D) face surface images were automatically produced by the Mirror Vectra CR10 program approximately 30 seconds after image capture (Figure 13). The fast image capture time allowed images to be obtained even when participants were unable to stay still. Images were exported as .obj 3D files, with corresponding .png texture files. There is no loss of image quality in these steps. Images could also be viewed in third party 3D image viewers (Figure 14).

Image quality was reviewed again by the operator (Vamsi Chinthapalli) during the landmarking step. A final check of image quality was performed by another experienced operator (Peter Hammond or Mike Suttie) prior to creation of the dense

surface model. In these steps, emphasis was placed on ensuring adequate coverage of the face surface and detection of artefacts in face shape. Three subject images were excluded in total, out of 869, due to image quality during the studies.



***Figure 13 The Mirror Vectra CR10 software program automatically displays each face surface image after image capture. The image can be rotated and zoomed in or out to look for any artefacts or missing areas of the face surface.***



***Figure 14 Another view of the face surface image in a third-party image viewer, which allows the image to be freely rotated, moved or magnified. All of these functions are helpful in checking image quality.***

## 2.4 OTHER DATA COLLECTION

Participants were asked directly for additional data after image capture:

- Ethnicity of themselves and their parents
- Handedness
- Any history of facial trauma, including fractures, surgery, scars and dentition
- Co-morbidities

Date of birth was taken from medical records to calculate exact age at the time of image capture. Participants also had height and weight measurements performed using a portable freestanding stadiometer and weighing scales. This was done in all cases by a doctor or nurse after heavy clothing and footwear was removed. Height and weight measurements were always within one week of image capture and in most cases was

taken just prior to image capture. Body mass index was calculated using the standard formula:

$$\text{Body mass index (BMI)} = \text{weight (kg)} / ((\text{height (m)} \times \text{height (m)})$$

Standard definitions were used to categorise people who were underweight (BMI<18.5), of normal weight (BMI 18.5 – 24.9), overweight (BMI 25 – 29.9) or obese (BMI >29.9). These definitions are used by the World Health Organization and the National Institutes of Health in the USA (Flegal et al., 2013).

Data were also obtained from the following sources:

- Patient medical records
- Nursing admission records (e.g. height and weight data)
- Other research databases (e.g. genotype status)

Other data for specific studies, such as MRI findings and anti-epileptic drug history, are described in the relevant sections.

## 2.5 LANDMARKING

One of the prerequisites for direct and indirect anthropometry is the identification of landmarks. In dense surface modelling, landmarks are used to co-register surface images for further comparison and analysis.

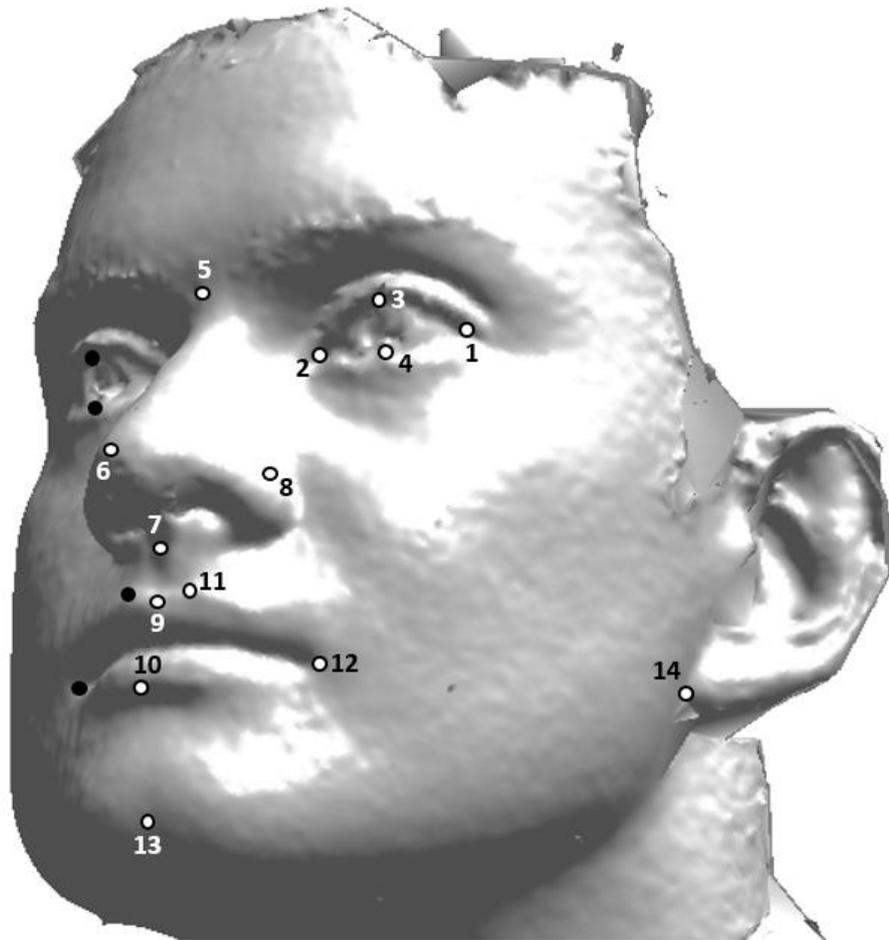
For this study, 22 well-defined soft tissue landmarks were chosen (Figure 15). They were chosen for ease of identification from visual inspection alone, high reproducibility, and to allow comparison with other previously landmarked images. All of these have previously been reproducible using stereophotogrammetry (Gwilliam et al., 2006; Plooij et al., 2009; Toma et al., 2009).

Landmarks are placed using the custom in-house Facemark software program (version 2.0, ShapeFind; UCL Institute of Child Health, London, UK). Surface images can be freely

rotated, displaced and magnified to aid placement of landmarks. The software also allows for wireframe views of the face surface. Wireframe views make it easier to identify certain facial contours.

One operator (Vamsi Chinthapalli) landmarked all images, for two reasons. Firstly, reproducibility of landmark placement on face surface images in 3D stereophotogrammetry is associated with the operator's level of familiarity with the imaging software (Gwilliam et al., 2006). Secondly, intra-operator reproducibility error is consistently smaller than inter-operator reproducibility error (Gwilliam et al., 2006; Plooij et al., 2009; Schaaf et al., 2010). Nevertheless, reproducibility of landmark placement was assessed for the current studies. An assessment of inter-operator reproducibility was initially performed to ensure that there was no significant systematic error between the operator here (Vamsi Chinthapalli) and another experienced operator (Peter Hammond; see Chapter 3). Assessment of inter-operator reproducibility also gives an indication of whether the method of stereophotogrammetry with landmarking could be expanded in subsequent clinical or research applications to have multiple operators.

Each landmark is described in more detail below and is shown in Figure 15.



*Figure 15 A typical face surface image, showing 14 of the 22 landmarks applied to the face (circles with black outline). The other 8 are paired landmarks on the right side of the face and are shown if visible (solid black circles). The landmarks are: 1. Exocanthion (paired) 2. Endocanthion (paired) 3. Palpebrale superius (paired) 4. Palpebrale inferius (paired) 5. Nasion 6. Pronasale 7. Subnasale 8. Alare nasi (paired) 9. Labiale superius 10. Labiale inferius 11. Cristae philtri (paired) 12. Cheilion (paired) 13. Gnathion 14. Otobasion inferius (paired)*

### 2.5.1 EXOCANTHION (PAIRED: EXOL AND EXOR)

Also termed the 'lateral canthus' or 'lateral palpebral commissure', this is the lateral junction of the upper and lower eyelid (Farkas, 1994; Hall et al., 2009). The angle here is usually more acute than for the endocanthion. It is readily identifiable on visual inspection due to different colour and contour between the skin and conjunctiva. This

landmark and the following three landmarks are difficult to measure using direct anthropometry due to patient discomfort on palpation.

#### 2.5.2 ENDOCANTHION (PAIRED: ENDOL AND ENDOR)

Also termed the 'medial canthus' or the 'medial palpebral commissure', this is the medial junction of the upper and lower eyelid (Farkas, 1994; Hall et al., 2009). Often the junction here extends towards the nose to form the lacus lacrimalis, in which tears collect before being drained by the lacrimal duct. In those cases, the actual junction is used. This landmark is also easily identified in a similar manner to exocanthion.

#### 2.5.3 PALPEBRALE SUPERIUS (PAIRED: PSUPL AND PSUPR)

This is the midpoint of the free margin of the upper eyelid. As the eyelid curves in three dimensions, it is difficult to use this definition in practice. Farkas uses the "highest point" of the free margin, but this requires the subject or image to be placed in the Frankfort horizontal plane, which may introduce further errors (Farkas, 1994).

Therefore, the midpoint can also be defined as the point on the free margin of the upper eyelid, which intersects with a line perpendicular to and bisecting the bicanthal line, as judged by visual inspection. The bicanthal line is the line connecting exocanthion and endocanthion. The position is affected by eyelid opening so care was taken to ensure subjects had their eyes fully open in images. This definition and method of placement have been used in a previous study assessing ptosis (Cox-Brinkman et al., 2007). For accuracy, the bicanthal line should be viewed perpendicularly by rotating the face laterally or there may be a parallax error in landmark placement.

#### 2.5.4 PALPEBRALE INFERIUS (PAIRED: PINFL AND PINFR)

This is defined in a similar fashion to palpebrale superius. It is the point on the free margin of the lower eyelid, which intersects with a line perpendicular to and bisecting the bicanthal line.

### 2.5.5 NASION (N)

The bony nasion is the point of intersection of the nasal bones and the frontal bone, at the superior end of the nasofrontal suture. The soft tissue nasion has been defined as the midline point overlying this (Hennekam et al., 2009) and is always superior to the endocanthion (Farkas, 1994). From visual inspection, it is approximately horizontal to the first crease overlying the upper eyelids, known as the supratarsal fold. This means that it is dependent on the degree of forward or backward tilt of the head (Gwilliam et al., 2006). It is not the deepest point along the nasofrontal angle, which is the subnasion, occurring inferior to the level of the supratarsal folds. Subnasion may be easier to identify in some individuals, but is also absent in those who have no depression between the nose and forehead. Nasion is incorrectly thought to be palpable due to it being mistaken for subnasion.

### 2.5.6 PRONASALE (PRN)

This is the most anterior point on the nose in the midline. It is more difficult to place in flat or bifid noses (Farkas, 1994). It is best identified when inspecting the face in profile view and again in frontal view.

### 2.5.7 SUBNASALE (SN)

This means 'below the nose' and has been recently defined as the base of the nose (Hennekam et al., 2009). It is the midpoint of the angle at which the columella and the surface of the upper lip meet (Farkas, 1994). The columella is the tissue that is between the nostrils when the nose is viewed from below, and forms the inferior border of the nasal septum. Like pronasale, this is seen best when the face is in profile view and then in frontal view. However, landmark placement is not affected by head tilt.

### 2.5.8 ALARI NASI (PAIRED: ALAL AND ALAR)

The alar curvature is the lateral wall of the nose and lateral border of the nostril. The alare (plural *alari*) is the most lateral point on the alar curvature. It is identified by viewing the face from inferiorly and from frontally. The lack of definite soft tissue features means there may be less reproducibility in its placement. The alari are used as they are the only non-midline landmarks of the nose, and therefore give information on nasal width and nostril shape. The alar crest is a related landmark that is the most lateral point of insertion of the alar curvature into the face. However, the alar crest does not give information on the anterior projection of the nostrils and so was not used.

### 2.5.9 LABIALE SUPERIUS (LS)

This is the centre of the upper vermilion border, judged on inspection to be equidistant from the cristae philtri and at the most inferior point on the curve of the vermilion border. The vermilion lip is the reddish part of the lips, due to thinner epidermis and visibility of underlying blood vessels. The vermilion border demarcates the vermilion lip from the adjacent keratinised skin of the upper lip (Carey et al., 2009). It is not necessarily in the midline of the face.

### 2.5.10 LABIALE INFERIUS (LI)

This is the centre of the lower vermilion border, which separates the lower vermilion lip from the keratinised skin of the rest of the lower lip.

### 2.5.11 CRISTAE PHILTRI (PAIRED: CHPL AND CHPR)

The shape of the upper lip has been described as 'Cupid's bow' in which the lateral portion of the vermilion border curves medially and superiorly until reaching two peaks, from which it continues medially and inferiorly to meet in the midline. Two pillars continue superiorly from the peaks to form the lateral boundary of the philtrum (Carey et al., 2009). The crista philtrum (plural *cristae philtri*) is defined as the

intersection of the upper vermilion border with the base of the pillar of the philtrum. The shape of the vermilion line and the elevation of the philtrum pillars both help to identify this landmark. However, one or both of these may be inconspicuous in some subjects, requiring image magnification.

#### 2.5.12 CHEILION (PAIRED: CHL AND CHR)

Also termed the 'labial commissure', this is the most lateral point of the oral cavity, or in common parlance, the corner of the mouth. It may be difficult to identify the edge of the oral cavity and the beginning of the skin crease of the angle of the mouth in some subjects and the vermilion border helps identification here.

#### 2.5.13 GNATHION (GN)

The bony gnathion is the most inferior median point on the lower border of the mandible. It has been traditionally used to denote the lowest point of the face. The soft tissue gnathion, also termed 'menton', is usually identified by palpation as the closest point on the surface of the face that corresponds to the bony gnathion (Farkas, 1994). From inspection, this is a difficult landmark to place. It is the only landmark for which palpation is traditionally preferred, but has been included to give an approximate representation of the vertical dimension of the face. The shape of a subject's chin is influenced by body weight and skin laxity. To help identification and reproducibility, gnathion is placed in the midline of the lower jaw at a point on the lower curve of the chin such that a tangent in the midsagittal plane would approximate 45 degrees to the horizontal.

#### 2.5.14 OTOBASION INFERIUS (PAIRED: OTOL AND OTOR)

This is the point of attachment of the ear lobe to the cheek (Farkas, 1994). It is important in determining facial width during analysis of face shape. It is also the easiest ear landmark to capture with the Vectra CR10 camera system. The posterior parts and superior parts of the external ear are more likely to be missed or obscured (for

example, by hair). Even when the whole ear is included, one group found that 3D surface images showed poorer resolution for the ear compared to other parts of the face (Gwilliam et al., 2006). Otobasion inferius was chosen instead of tragon or subaurale (the most inferior point of the earlobe), due to greater objectivity in placement. For example, attached earlobes make it impossible to place subaurale accurately.

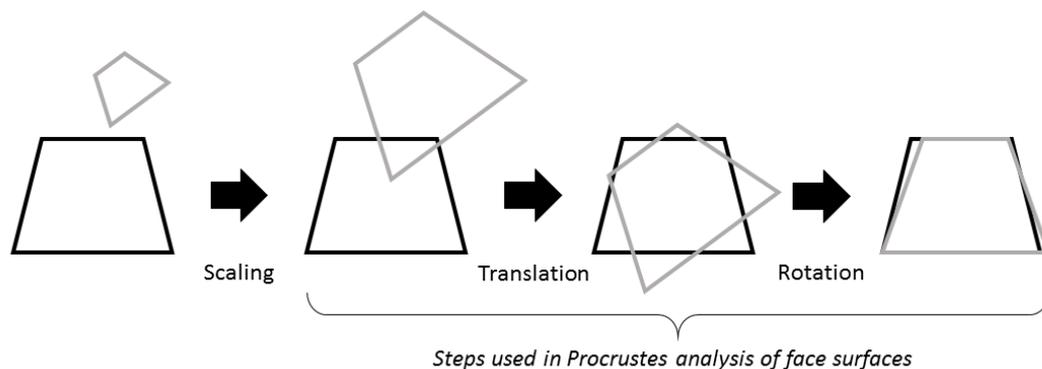
## 2.6 DENSE SURFACE MODELLING

At this stage, it should be emphasised that the 3D face surface comprises over 20,000 point co-ordinates, of which 22 have been manually designated as landmarks in the 'landmarking' step in the previous section. Figure 18 below shows a wireframe view of a face surface showing these point co-ordinates. The next steps involved the creation of a dense surface model (DSM). The steps can be broken down into the following:

1. Procrustes analysis
2. Thin plate spline
3. Dense correspondence creation
4. Reverse thin plate spline
5. Principal component (PC) analysis
6. Face Shape Difference (FSD) calculation

Previous studies have used the methods described here for dense surface model creation (Bhuiyan et al., 2006; Cox-Brinkman et al., 2007; de Souza et al., 2013; Hammond, 2007; Hammond et al., 2012a; P. Hammond et al., 2008; Hammond et al., 2005, 2004; Heulens et al., 2013; Suttie et al., 2013). An in-house software program called DSMEexplorer was used (version 2.1, ShapeFind; UCL Institute of Child Health, London, UK; made available by Peter Hammond) as in previous studies. The input data required are the raw object files for every subject's face image and the list of point co-

ordinates of all 22 landmarks for each image. A number of steps are used to generate the model and these are also described mathematically elsewhere (Hutton et al., 2003, 2001). The following steps of generalised Procrustes analysis, thin-plate spline warping, dense correspondence creation and principal component analysis are all performed using the software program.

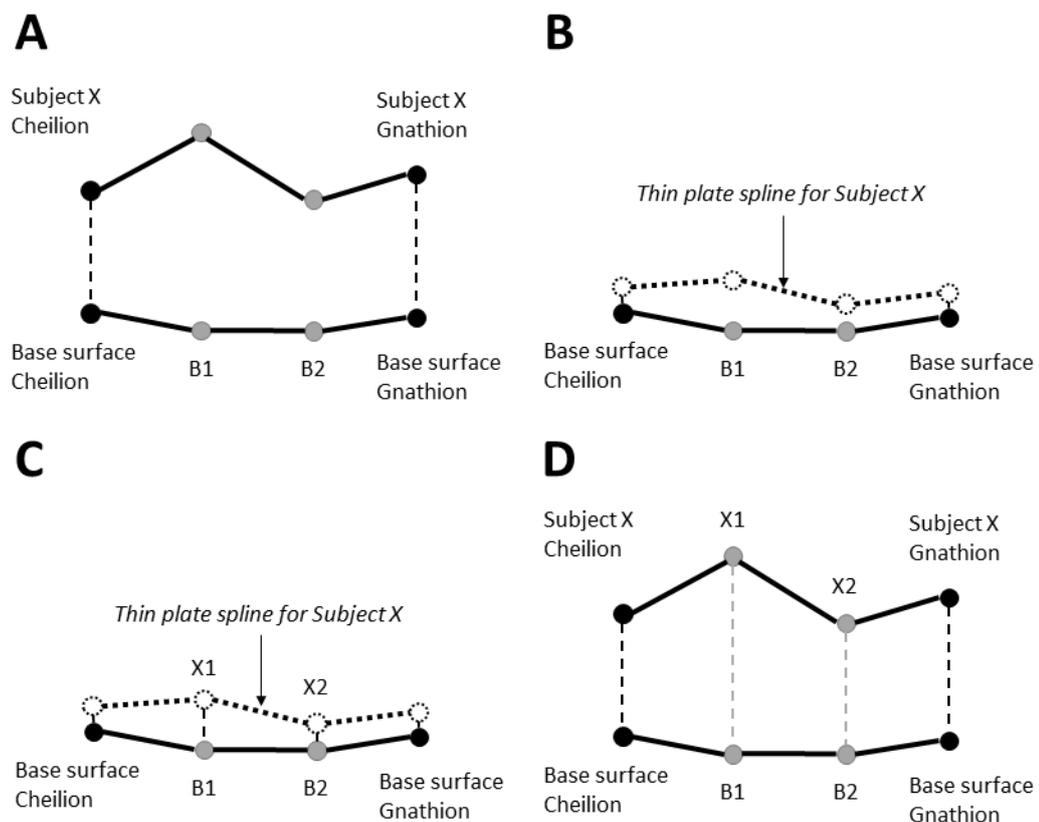


**Figure 16** An example of Procrustes analysis with two trapeziums (outlined in grey and black). Partial Procrustes analysis uses rotation and translation to superimpose shapes on one another. Full Procrustes analysis also uses scaling, but in this study, face size could be an important factor so no scaling was employed. Procrustes analysis can be employed with any number of landmarks, not just four as in the example above. When Procrustes analysis is used with more than two objects, it is called generalised Procrustes analysis.

First, the images have to be co-registered or aligned using generalised Procrustes analysis. It is named after Procrustes, a character in Greek mythology who tortured victims by stretching their bodies or amputating their legs to make them fit on his bed. The analysis is a mathematical procedure to superimpose shapes as closely as possible on to one another (Figure 16). The shapes in this case were whole face surfaces. Superimposition is by means of rotating and translating shapes. Scaling is employed in some forms of Procrustes analysis but was not used in this thesis because face size information would be lost when all face surfaces are scaled to the same size. Face size may be affected, as well as face shape, by structural variants, body mass index and developmental brain lesions. The closest fit for all face surfaces is done using the 'least-

squares' method of alignment, in which the sum of the square distances between each landmark in a pair of face surfaces is minimised. After Procrustes analysis, the actual *shapes* of the face surfaces have not been altered.

In the next step, the shapes are warped on to a base surface shape so that all surface points can be labelled and included in analysis (Figure 17). This base surface can be any of the face surfaces in the group of faces being analysed and the choice of base surface makes no difference to later analysis (Hutton et al., 2001). The warping is performed using another mathematical technique known as thin-plate spline.



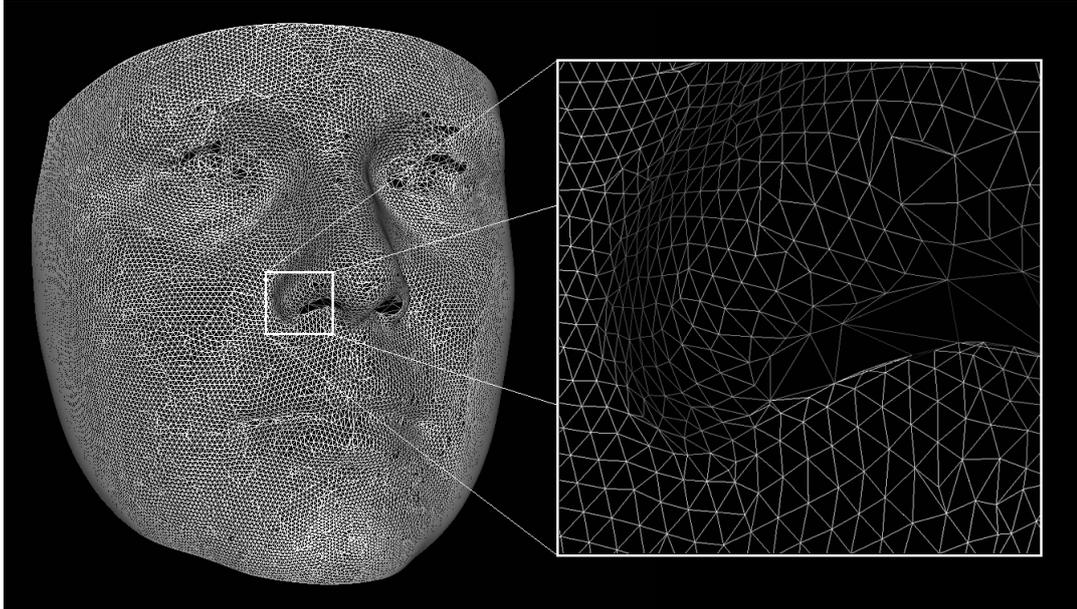
**Figure 17** Panel A shows an example of a simplified cross-section of part of two face surfaces, a base surface and Subject X's face surface. Two landmarks, cheilion and gnathion, are shown with two surface points in between them. In reality there are hundreds of surface points in between these landmarks. Panel B shows thin plate spline being applied to the Subject X face surface, which is a mathematical function to ensure maximum closeness of fit of each surface point with the minimum deformation required. Panel C shows that once the surfaces are closely apposed after thin plate spline, a 'dense correspondence' can be created with labelled surface points interpolated between the landmarks for the Subject X face surface. In this case, surface points B1 and B2 on the base surface are matched to

*equivalent points on the Subject X face surface, which can be termed X1 and X2. Panel D shows the reversal of thin plate spline to recreate the original face surface for Subject X. However, now the surface points all correspond to individual surface points on the base surface.*

Thin-plate spline can alter shapes such that fixed points (in this case the landmarks) are perfectly aligned for every face surface and all of the other surface points are as 'smoothly' transformed on to the base surface as possible. In mathematical terms, the smoothness is due to the use of a function to balance the closeness of fit of all surface points with the minimum deformation (measured as the displacement of each point) needed. Note that thin plate spline is only used so that every face surface point on every face can correspond to an equivalent point, just as the landmarks are. This 'dense correspondence' allows sophisticated analysis of face shape not possible with landmarks alone. For example, after landmarking, the surface point of gnathion in one subject corresponds to a surface point labelled gnathion in every other subject. Dense correspondence means, for example, that the point just superior to gnathion corresponds to the point just superior to gnathion in all face surfaces of subjects. This is also true of each and every one of the over 20,000 surface points. Without dense correspondence, only landmarks themselves would be compared in shape analysis and variations in face shape in areas outside landmarks would be ignored, such as on the cheeks or forehead. Areas between landmarks, such as the shape of the nasal bridge between nasion and pronasale, would be ignored too. Thin plate spline, in turn, is necessary for the software program to create an accurate dense correspondence. Once the dense correspondence has been made for every image, the images undergo a reverse thin-plate spline warp to return them to their original shape. Now the dense correspondence has created approximately 20,000 'landmarks' that are common between the faces and can be used for comparison (Figure 18).

Each face image may have surface points covering extraneous areas such as the neck, that may or may not be included in other images. These areas are removed by ignoring

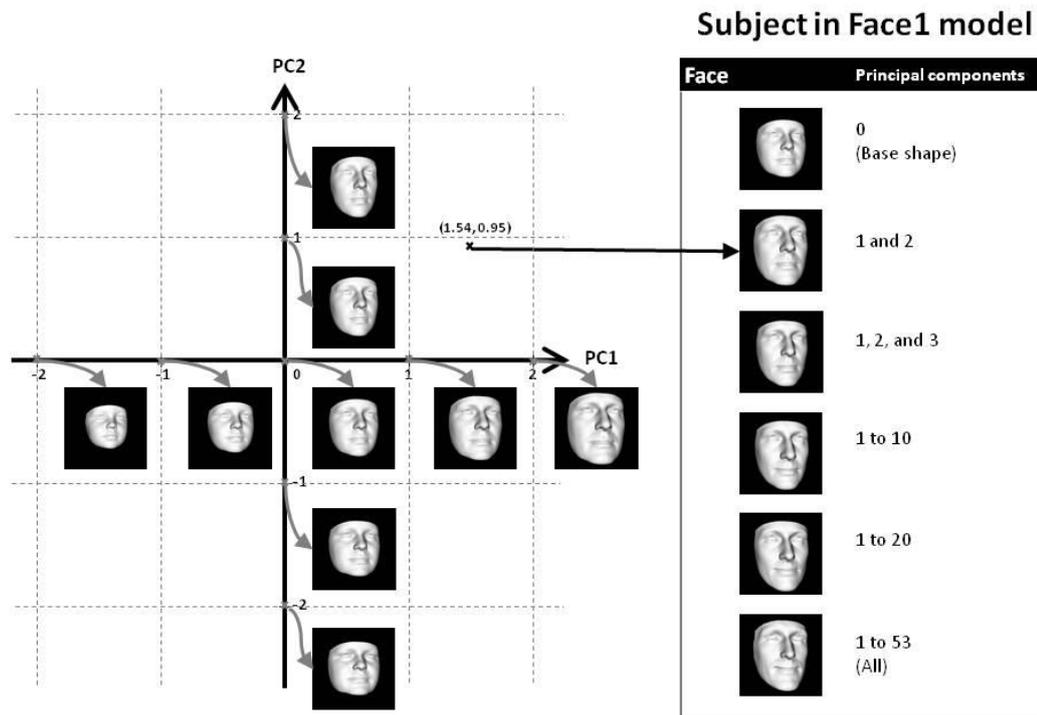
all points that do not form a dense correspondence within a certain distance (20mm). Thus, after the thin-plate spline and dense correspondence steps, all face images cover the same regions of the face.



**Figure 18** A wireframe view of the mean face surface in one of the studies. This shows all of the dense correspondence points created after thin plate spline interpolation, connected to each other with straight lines. The right nostril has been magnified to show the points more clearly. Each correspondence point can be compared in all of the subjects within the dense surface model (DSM).

To compare the different shapes, principal component (PC) analysis was used. There are three reasons for this. Firstly, PC analysis allows a manageable dataset to be created, which can be utilised on personal computers. Secondly, the principal components illustrate specific characteristics of changes in face shape and size. For example, variation in one PC was shown to help classify a set of face surfaces as having Noonan syndrome or not having the syndrome (Hammond et al., 2004). Thirdly, PC analysis can provide a single metric to assess how atypical the shape of a face is – described later. A PC is a particular scalable transformation. PC analysis describes the overall transformation needed to obtain one shape from another using the smallest number of PCs to describe the greatest amount of variation between shapes. The first PC is computed to have the largest possible variance and so it is always the one that

describes as much of the variation as possible in one transformation. The PC is scalable in that a numerical coefficient can describe the extent to which any shape is transformed along the 'axis' of the PC (Figure 19). The second PC is the transformation that adjusts for the greatest amount of the remaining variation and is orthogonal, that is non-overlapping, to the first PC. This is continued until 99.0% of all face shape variation was explained by a number of successive PCs. In other words, a weighted linear sum of all PCs with their coefficients can describe each face surface shape, instead of approximately 20,000 point co-ordinates (Figure 19). The figure of 99% is a compromise between creating a manageable dataset and minimising face shape data loss. It is greater than in the only studies of face shape using principal components by other groups, which have used thresholds of 75-87% (Hennessy et al., 2010, 2007, 2002; Prasad et al., 2015). All of these studies investigated one particular phenotype so detection of all variation was less important. In the present study, it was important to identify and detect any unrecognised facial dysmorphism.



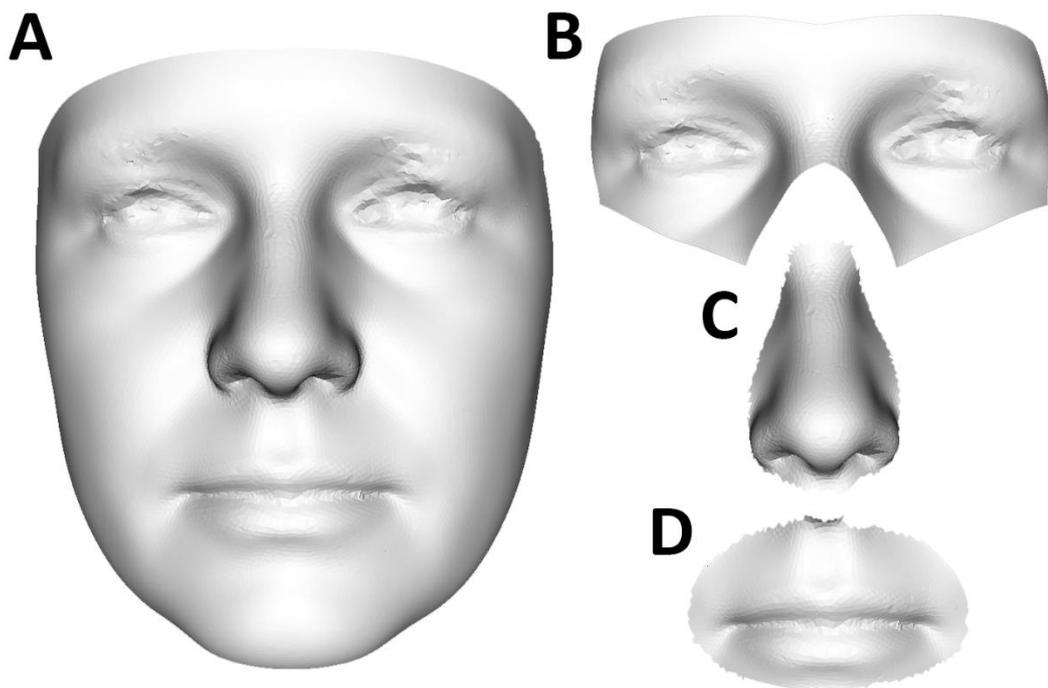
**Figure 19** Each principal component can be thought of as one dimension of shape variation along which a face surface can be transformed. The units or coefficients of each principal component represent the degree of transformation from the mean face surface. Only the first two principal components (PC1 and PC2) are shown here in two dimensions, with PC1 along the x-axis and PC2 along the y-axis. In this dense surface model (DSM), the base surface is shown at (0,0). Increasing the value of PC1 increases facial size, which can be seen in the row of five images from (-2,0) to (2,0). Increasing PC2 elongates and narrows the facial shape, seen between (0,-2) and (0,2). Every subject's face can be described by a point in this space, such as (1.54,0.95) for the above subject. However, the two principal components only capture 84.5% of facial variation in this dense surface model, Face1, and 53 principal components are needed to capture over 99% of variation. This increased detail for the face shape can be seen in the panel on the right.

A number of different groups of subjects were studied in my work and they are described in detail in Section 2.2.5. For each group of subjects, a separate DSM was created. The DSM covered the whole face (Figure 20 panel A) and was called Face1 for Group 1 subjects, Face2 for Group 2 subjects, and so on. Another DSM for Group 1 subjects, called FaceR, was created to study reproducibility (Chapter 3). In total, eight DSMs were created for face shape and called Face1-Face7 and FaceR.

The same groups were also used to create DSMs of just particular parts of the face.

These parts were 'feature-rich' areas with multiple landmarks and contour changes: the

periorbital region only (Figure 20 panel B), the perinasal region only (Figure 20 panel C) and the perioral region only (Figure 20 panel D). These DSMs of parts of the face were used to explore the extent of regional face shape change and to see if DSMs of a part of the face were as discriminating and useful as a DSM of the whole face. Initially, software limitations meant that only DSMs of the periorbital region (Eyes1-Eyes6, EyesR) and the perinasal region (Nose1-Nose6, NoseR) were made. After a software upgrade, I was able to create DSMs of the perioral region for later studies (Mouth2-7). The DSMs of parts of the face were created in the same way as the DSM of the whole face except that they were trimmed to the region of interest only (Figure 20).



**Figure 20** This shows the four different DSMs used in analyses throughout this study. The primary DSM was for the whole face (A). Three further DSMs investigated the usefulness of using one part of the face: the periorbital region only (B), the perinasal region only (C) or the perioral region only (D).

Usually the more surfaces that are included in a DSM, the more PCs are needed to describe 99.0% of the face shape. In my studies, different numbers of principal components were needed to explain at least 99.0% of the face shape variation. DSMs are created independently from each other and may have a different number of PCs.

The action of each PC will also be different between DSMs. This means that PCs cannot be directly compared across DSMs. Hence in two validation studies here (which use either Group 2 and Group 4), the training set was included in the group as well as the validation set – but of course the validation set was assessed separately with no overlapping subjects in both the training and validation set. A summary of the different DSMs for each group of subjects and the number of PCs used is shown in Table 4.

Group	Chapter	Whole face		Periorbital		Perinasal		Perioral	
R	3	FaceR	53	EyesR	98	NoseR	46	N/A	
1	3,4	Face1	53	Eyes1	98	Nose1	46	N/A	
2	4,5	Face2	81	Eyes2	172	Nose2	70	Mouth2	79
3	3,6	Face3	103	Eyes3	172	Nose3	70	Mouth3	79
4	6	Face4	96	Eyes4	229	Nose4	87	Mouth4	72
5	6	Face5	96	Eyes5	229	Nose5	87	Mouth5	72
6	7	Face6	97	Eyes6	239	Nose6	79	Mouth6	92
7	8	Face7	85	N/A		N/A		N/A	

**Table 4 Summary of the DSMs made for each group of subjects in the following studies here. Group R is the same as Group 1 except with some landmarks replaced in some subjects as part of a reproducibility study (Chapter 4). Groups 1 to 7 are the groups of subjects described in Section 2.2.5. Groups 1 and 2 were used in the study of pathogenic structural variants in Chapters 4 and 5. Groups 3-5 were used in the study of BMI and age in Chapter 6. Group 6 was used in the study of developmental brain lesions in Chapter 7. Group 7 was used in the study of people with HNF1B mutations in Chapter 8. For each group, one DSM was made of the whole face and in some cases, more DSMs of other parts of the face. For each of these DSMs, the number of principal components (PCs) used is shown next to them.**

## 2.7 STATISTICAL ANALYSIS

### 2.7.1 FACE SHAPE DIFFERENCE (FSD)

Principal components provide a measure of the degree of facial surface transformation needed to reproduce the subject's surface relative to an average matched face in a particular DSM. Differences in a subset of principal components have been used previously to discriminate between controls and people with a particular condition graphically. As part of the thesis, I sought to explore face shape in people with different underlying structural variants. I was not seeking to compare one homogeneous group

to a control group. Therefore, I created a measurement of overall facial dysmorphism, which was checked for technical validity by Peter Hammond.

Within each DSM, I compared each subject face to an average control face. This control face was created using the mean of each and every principal component coefficient in a set of the 30 closest age-matched sex-matched ethnicity-matched reference control faces.

Then, the square root of the sum of the squared differences between every principal component coefficient in the subject and the mean of the control set was calculated to give a 'Face Shape Difference' (FSD). The FSD value is an indicator of the difference in PCs in a subject relative to the matched control face surfaces. The FSD for a subject can be expressed algebraically as follows:

$$\text{FSD}(x) = \sqrt{(\sum(x_{PC1} - m_{PC1})^2 + (x_{PC2} - m_{PC2})^2 + (x_{PC3} - m_{PC3})^2 + \dots)}$$

Or more concisely as:

$$\text{FSD}(x,m) = \sqrt{\sum_{i=1}^n (x_i - m_i)^2} \text{ for } x \in M$$

where  $i$  indexes the  $n$  principal components ( $PC1, PC2, PC3$ , etc.) capturing 99% shape variation in dense surface model  $M$  and  $x$  is an arbitrary subject's face which is compared to  $m$ , its matched mean from the control set.

The square root of the sum of the squared differences is a statistical tool that is used to combine the difference between unrelated variables, as is the case with PCs.

During matching, if a subject was not of the same ethnicity, he or she was excluded from further analysis. There were sufficient controls to generate mean PCs between the ages of 2.6 to 50 years for men and 2.7 to 53 years for women. The maximum discrepancy in age between subjects and the matched control set mean age was kept at

9 months for children and 3 years for adults. The average discrepancy in age was 0.44% for male subjects and 0.74% for female subjects.

### 2.7.2 REFLECTED FACE SHAPE DIFFERENCE

To test asymmetry, an additional face image was created for every subject using the mirror image of the original face surface with landmarks. The landmarks were relabelled from 'left' to 'right' and vice versa for all paired landmarks and no change was made to midline landmarks to avoid introducing any error from repeat landmarking. After the DSM was created including these mirror images for subjects, I calculated the distance between each face and its mirror image in terms of the square root of the sum of squared differences of every principal component. Algebraically, the formula is:

$$\text{FSD}(x,r) = \sqrt{\sum_{i=1}^n (x_i - r_i)^2} \text{ for } x \in M$$

where  $i$  indexes the  $n$  principal components ( $PC1$ ,  $PC2$ ,  $PC3$ , etc.) capturing 99% shape variation in dense surface model  $M$  and  $x$  is an arbitrary subject's face which is compared to  $r$ , its reflected face surface image.

This distance was termed the 'reflected face shape difference' (reflected FSD) and a similar measure has been used before to detect facial asymmetry in autism spectrum disorders (Hammond et al., 2008). Like FSD, reflected FSD only measures the extent of any difference in the face images.

One can see that for reflected FSD, control subjects are not necessary as a subject's face is compared only with its reflection.

### 2.7.3 POWER CALCULATION

The following power calculations were for two of my studies assessing FSD in people with pathogenic SVs and in people with unilateral developmental brain lesions. The

tests were to detect a two-sided difference between the means of two sample groups and calculated using a statistical website (<http://powerandsamplesize.com>) using the following formula, in which the two sample groups are designated A and B, with  $n_A$  people in A,  $n_B$  people in B, and the sample size ratio or matching ratio described as  $n_A/n_B$ :

$$n_A = \kappa n_B \text{ and } n_B = \left(1 + \frac{1}{\kappa}\right) \left(\sigma \frac{z_{1-\alpha/2} + z_{1-\beta}}{\mu_A - \mu_B}\right)^2$$

$$1 - \beta = \Phi(z - z_{1-\alpha/2}) + \Phi(-z - z_{1-\alpha/2}) \quad , \quad z = \frac{\mu_A - \mu_B}{\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

where

$\kappa = n_A/n_B$  is the matching ratio

$\sigma$  is standard deviation

$\Phi$  is the [standard Normal distribution function](#)

$\Phi^{-1}$  is the [standard Normal quantile function](#)

$\alpha$  is Type I error

$\beta$  is Type II error, meaning  $1 - \beta$  is power

There has been no known previous use of FSD before and it was unclear, *a priori*, how different it would be in different samples of people with epilepsy. However, there have been studies looking at differences in the underlying PCs that make up FSD. A study of 22q11 deletion syndrome shows a difference in three of the first four PCs and in one of these PCs, the mean difference is approximately 3 units with a standard deviation of 5.6 units (derived from Figure 11A in Hammond et al., 2004). Assuming that the 22q11 deletion syndrome is a more easily recognisable dysmorphic syndrome than most and that the mean difference in a group of all pathogenic SVs would be much less, I chose a conservative estimate of mean difference of 1 unit with the same standard deviation, expected to be found in three PCs. Therefore, the mean difference in FSD and the sample size can be worked out as:

Assumed total number of PCs in the DSM = 50 (Hammond et al., 2004)

Number of PCs with 1 unit difference between affected and control subjects= 3

Number of PCs with 0 units difference between affected and control subjects = 47

$$\text{Mean difference in FSD} = \sqrt{(1 \times 3) + (0 \times 47)} = 1.7$$

$\mu_a = 1.7$  Mean FSD increase in affected subjects

$\mu_b = 0.0$  Mean FSD increase in control subjects

$\sigma = 5.6$  Standard deviation of affected subjects

$\alpha = 0.05$  Type 1 error threshold

$1-\beta = 0.8$  Type 2 error threshold

$\kappa = 0.2$  Sampling ratio of 1 affected subject per 5 unaffected subjects used

From the formula described at the start of this section, the minimum required sample size is calculated as: 102 unaffected subjects and 20 affected subjects.

There is one study that quantifies differences in PCs due to asymmetry (Hammond et al., 2008). This study found that in comparing those with autism to controls, the difference in face shape using an “asymmetry index”, another statistical measure derived from DSMs and PCs, was approximately 0.67 (6.77 in autism and 6.10 in controls) with a variance of 2.84 (hence standard deviation of  $\sqrt{2.84} = 1.69$ ).

Assuming that autism and developmental brain lesions may be associated with a similar degree of facial asymmetry, the required sample size can be calculated from:

$\mu_a = 6.77$  Mean asymmetry index in affected subjects

$\mu_b = 6.10$  Mean asymmetry index in unaffected subjects

$\sigma = 1.69$  Standard deviation of affected subjects

$\alpha = 0.05$  Type 1 error threshold

$1-\beta = 0.8$  Type 2 error threshold

$\kappa = 0.2$  Sampling ratio of 1 affected subject per 5 unaffected subjects used

From the formula described at the start of this section, the minimum required sample size is calculated as: 240 unaffected subjects and 60 affected subjects.

#### 2.7.4 STATISTICAL TESTS USED WITH FSD

Statistical tests used in each analysis are described in detail in subsequent chapters. An  $\alpha$  value of 0.05 was used in all tests unless otherwise specified. In summary, normality of data was assessed with the Shapiro-Wilks test. For non-parametric data, the Wilcoxon Mann-Whitney test and Kruskal-Wallis tests were used with the null hypothesis that FSD was equal in both groups of subjects. For parametric data, the paired or unpaired t-test and the ANOVA tests were used. Spearman's rank correlation coefficient or Pearson's correlation coefficient were used to compare associations between data. Analysis was conducted using SPSS software (versions 19-23, SPSS Inc; Chicago, IL, USA).

#### 2.8 ETHICS APPROVAL

This study was approved by the Joint Research Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, University College London as part of 'A Population Based Genetic Study of Epilepsy' (Study Number 00/N081, Chief Investigator: Sanjay Sisodiya). The relevant amendment was approved on 26<sup>th</sup> March 2009. The study was sponsored by University College Hospitals National Health Service Foundation Trust and participants were recruited from two locations within the Trust:

- National Hospital for Neurology and Neurosurgery, Queen Square, London
- Chalfont Centre for Epilepsy, Chesham Lane, Chalfont St Peter

Reference control subjects had been previously recruited by Peter Hammond after written, informed consent and approval by University College London Hospital Research Ethics Committee.

Recruitment from European centres is described above.

#### 2.9 DATA PROTECTION

All image data was stored in coded form on an encrypted drive of a personal computer used with the stereophotogrammetric camera. The personal computer was stored in a room that was locked with a key and also had a combination door lock.

A backup portable hard drive and desktop hard drive also contained coded image data. Both of these were also encrypted. They were stored in a room at a different physical site, which was accessed using a swipe card and a combination door lock.

Access to the rooms was limited to researchers in the same department and passwords for the encrypted drive were known only to four research staff who were directly recruiting participants and performing image capture. Codes for participant data were stored in an encrypted spreadsheet, in a restricted folder on University College London Hospitals NHS Foundation Trust servers.

## CHAPTER 3: VALIDATION OF STEREOPHOTOGRAMMETRY AND DENSE SURFACE MODELLING

### 3.1 INTRODUCTION

Stereophotogrammetry is a method of measurement and as such, it must be scientifically validated before being used in clinical research. Validation is even more important for stereophotogrammetry because it is a relatively novel technique and it contains a number of steps that can each introduce error.

In 1996, Farkas, the leading researcher using direct anthropometry, stated that there was a need to prove the reliability of 3D surface scanning methods (Farkas and Deutsch, 1996). Some earlier research had noted high accuracy to within 0.3mm but not in living subjects or systematically (Beard and Burke, 1967; Thomas et al., 1989). The first study in which 3D indirect anthropometry was compared to direct anthropometry was performed using laser surface scanning of 30 young adults in 1995 (Aung et al., 1995). The authors found that they could identify 37 out of 41 standard anthropometric facial landmarks from the scan whereas four needed to be marked on the face beforehand. In terms of accuracy, just over half of all linear measurements between landmarks (42 of 83) were deemed unreliable because they showed a mean difference of >2mm between the two methods – they also tended to be the longest measurements although the authors do not comment upon that. The least reliable measurements used landmarks at the periphery of the face such as the tragus, glabella and gnathion. The errors were partly due to the vertical scanning movement of the laser beam being less suited for horizontal plane accuracy, the 15 second image acquisition time and the difficulty of identifying landmarks on a laser scan, which lacks photographic information of the surface. Despite these limitations, for over a third of measurements (28 of 83), the mean error was less than 1.5mm. Since then,

stereophotogrammetry has entered widespread clinical practice and reproducibility has been assessed using this method of image capture.

### 3.1.1 VALIDATION STUDIES OF CURRENT STEREOPHOTOGRAMMETRY CAMERAS

Over the last 13 years, there have been many other studies of accuracy and reproducibility using different stereophotogrammetry systems and configurations (Table 5). Accuracy refers to how close a measurement is to the true value – and in the majority of these studies, the ‘gold standard’ for measurements of the true value were performed using direct anthropometry. Reproducibility or precision is the degree to which repeated measurements are the same, and were usually performed for the same operator (intra-operator) and between operators (inter-operator). I will look at these studies in more detail regarding the overall accuracy and reproducibility, and also for the following individual steps in the image analysis procedure:

1. Image capture and registration
2. Head position
3. Face expression
4. Face dysmorphism
5. Landmarking
6. Dense surface model creation

Three current camera manufacturers are represented in the scientific literature and they are 3dMD (3dMD; London, UK), Canfield (Fairfield, NJ, USA: Canfield Scientific Inc) and Di3D (Dimensional Imaging; Glasgow, UK). Their models all use passive stereophotogrammetry but with a synchronised camera flash. According to the three manufacturers’ camera specifications, image capture speed is 1-3.5ms, accuracy is within 0.2mm, and all create a 3D mesh framework with a colour surface image (Tzou et al., 2014). Due to the similarities between the cameras, studies using all of them for face shape measurement were analysed (Table 5 and Table 6).

It should be noted that seven studies were of mannequin heads or casts of the face. This may underestimate error as there is no movement, no change in expression, and they lack soft tissue imaging characteristics. Loss of image detail due to soft tissue reflection is noted in one study (Ghoddousi et al., 2007). By contrast, direct measurements on human faces may not be a 'gold standard' in certain cases and mannequins/casts may be superior. Facial soft tissue may show slight deformation during direct anthropometry using callipers or measuring tape. For example, two studies noted the difficulty of accurately measuring the inner and outer canthus of the eye with callipers due to delicate and sensitive structures around them (Ghoddousi et al., 2007; Heike et al., 2009). Thus it is useful to include studies of both human faces and of casts or mannequins of the human face, because each type has its own limitations.

### 3.1.2 STUDIES OF ACCURACY OF STEREOPHOTOGRAMMETRY

Twelve studies compare the accuracy of 3D stereophotogrammetry to direct anthropometry as a gold standard (Chen et al., 2015; Dindaroğlu et al., 2015; Fourie et al., 2011; Ghoddousi et al., 2007; Heike et al., 2009; Kook et al., 2014; Lee et al., 2004; Lübbers et al., 2010; Schaaf et al., 2010; Weinberg et al., 2006; Winder et al., 2008; Wong et al., 2008). Two others compare it to specialised direct co-ordinate measurement machines and one compares it to CT scans (Ayoub et al., 2003; Khambay et al., 2008; Metzger et al., 2013). Another five studies look at reproducibility only (Aldridge et al., 2005; de Menezes et al., 2010; Gwilliam et al., 2006; Othman et al., 2013; Plooij et al., 2009). They are all listed in Table 5 and Table 6. One other study comparing stereophotogrammetry to moiré photography was excluded because the authors looked at agreement between both methods but not accuracy or precision (Artopoulos et al., 2014).

The studies all show that stereophotogrammetry using any of the cameras has an accuracy of within 1mm compared to direct anthropometry, although often worse than

that stated by the manufacturers. Therefore this would suggest that stereophotogrammetry is superior to 2D indirect anthropometry, laser scanning and moiré photography in terms of accuracy.

### 3.1.3 STUDIES OF REPRODUCIBILITY OF STEREOGRAMMETRY

Nine studies assessed the absolute reproducibility error (Aldridge et al., 2005; Ayoub et al., 2003; de Menezes et al., 2010; Gwilliam et al., 2006; Plooij et al., 2009; Schaaf et al., 2010; Weinberg et al., 2006; Winder et al., 2008; Wong et al., 2008) and two other studies only looked at intraclass correlation coefficients for reproducibility (Dindaroğlu et al., 2015; Othman et al., 2013).

All studies showed that intra-operator and inter-operator reproducibility was within 1mm with the exception of two studies. One of these studies was in an obsolete camera model known to have less surface point resolution than newer cameras even in 2004 (Gwilliam et al., 2006; Hammond et al., 2004). The other study used only two landmarks, left and right frontotemporal, that were each used to measure two diagonal distances across the skull vault with no corresponding formal landmark on the opposite side of the skull (Schaaf et al., 2010). Instead the authors took the end point of their distances as the “contralateral occipital area close to the lambdoid suture”. Using only one clearly defined landmark to measure a linear distance (in this case, the oblique length of the cranium in infants) would appear to be a less accurate method than using a start and end landmark, but this has not been studied. Both landmarks are on the lateral forehead and not commonly used in anthropometric studies. The landmarks were also placed on images of infants’ heads covered by a tight-fitting cap, further reducing accuracy as the forehead itself was not visible.

All studies that compared intra-operator reproducibility to inter-operator reproducibility found that intra-operator reproducibility was greater, except for one study (de Menezes et al., 2010). This study compared intra-operator reproducibility for

measurements from two different images of each subject's face to inter-operator reproducibility for the same image of each subject's face. Therefore, the intra-operator reproducibility error also includes the errors from a second image acquisition.

Reproducibility error has been poorly studied in direct anthropometry. A number of studies comment on how errors can be introduced by poor landmark identification, poor instruments and poor measurement technique (Allanson et al., 1993; Hunter, 1996; Jamison and Ward, 1993). One author comments on the need for further studies of these errors (Allanson, 1997). It is likely that reproducibility error is smaller for stereophotogrammetry than for direct anthropometry although the two methods have not been compared.

Study	Year	Camera model	Comparison	Face	Age	Area of face	Accuracy (mm)	Reproducibility (mm)
Ayoub et al	2003	Proprietary research system	CM	M	I	Whole face	0.42	0.20
Lee et al	2004	Genex	DA	M	A	Whole face	1.11	
Aldridge et al	2005	3dMD	Nil	H	A, C	Whole face		0.83
Weinberg et al	2006	3dMD, Genex Facecam 250	DA	M	A	Whole face	<1.00	<0.60
Gwilliam et al	2006	DSP400	Nil	H	A, C	Whole face		50% <1.0 (IA), 8% <1.0 (IE)
Ghoddousi et al	2007	(Canfield Vectra CR10)	DA, 2DP	H	A	Whole face	0.23	
Khambay et al	2008	Di3D System	CM	M	A	Whole face	0.20	
Winder et al	2008	Di3D System	DA	M	A	Whole face	0.62	0.06
Wong et al	2008	3dMDface System	DA	H	A	Whole face	0.9	0.8
Heike et al	2009	3dMDface System	DA	H	A, C	Whole face	<1.00	
Plooiij et al	2009	3dMDface System	Nil	H	A	Whole face		0.26 (IA), 0.32 (IE)
Lubbers et al	2010	3dMD	DA	M	A	Whole face	0.41	
Schaaf et al	2010	3dMD Cranial System	DA	H	I	Cranium		1.1-1.3 (IA), 2.6-2.8 (IE)
de Menezes et al	2010	Canfield Vectra CR10	Nil	H	A	Whole face		0.53 (IA), 0.29 (IE)
Fourie et al	2011	Di3D System	DA, LS, CT	C	A	Whole face	0.86	
Metzger et al	2013	3dMD	CT	H	A	Whole face	<i>2/3 of landmarks show no sig difference</i>	
Othman et al	2013	Canfield Vectra CR10	Nil	H	A	Whole face		<i>ICC = 0.68-0.97 (IA)</i>
Kook et al	2014	Di3D System	DA, CT, LS	M	A	Whole face	0.32-0.85	
Chen et al	2015	3dMD Cranial System	DA	H	A	Ear auricle	0.27	
Dindaroglu et al	2015	3dMDflex System	DA, 2DP	H	A	Whole face	0.21	<i>ICC = 0.94 (IA), 0.94 (IE)</i>

**Table 5 Studies of accuracy and precision of stereophotogrammetry, where accuracy is in comparison to another measurement method in the Comparison column and reproducibility is for repeated measurements using stereophotogrammetry alone. If known, IA is used for intra-operator reproducibility and IE for inter-operator reproducibility. Measurements with ICC are only for the intraclass correlation coefficient and were not given as absolute errors in the studies. The Comparison column shows other methods of measurement used: 2DP = 2D photography, CM = Co-ordinate measuring machine, CT = Computed tomography, DA = Direct anthropometry, LS = Laser scanning. The face column shows the subjects used: C = Cadaver face, M = mannequin or cast of face, H = Live human subject. The Age column lists the age of subjects: I = Infants, C = Children, A = Adults.**

<b>Study</b>	<b>Year</b>	<b>Measurements*</b>	<b>Least accurate landmarks</b>	<b>Landmarked^</b>	<b>Notes</b>
Ayoub et al	2003	Co-ordinates	Pronasale, alare nasi	Yes	Infants with cleft lip
Lee et al	2004	Linear	Zygoma	Yes	Assessed at 4 different angles
Aldridge et al	2005	Linear	Nasion, tragion	No	Children with Down syndrome/craniosynostosis
Weinberg et al	2006	Linear	Nasion, gnathion, exocanthion, alare nasi	Yes	
Gwilliam et al	2006	Co-ordinates	Alar crest, nasion, gonion, menton	No	Measures landmarking precision
Ghoddousi et al	2007	Linear	Intercanthal width, alar length, upper lip	Yes	
Khambay et al	2008	Co-ordinates	Nostrils	No	Measures landmarking accuracy
Winder et al	2008	Linear	Cheilion, glabella, gnathion	Yes	
Wong et al	2008	Linear	Endocanthion, tragus, gnathion	No	
Heike et al	2009	Linear	Columella width, upper face height, mouth width	Yes	22q11.2 deletion syndrome
Plooij et al	2009	Co-ordinates	Alare, tragion	No	Measures landmarking precision
Lubbers et al	2010	Linear	Not assessed	Yes	0.10mm operator & 0.11mm calibration error
Schaaf et al	2010	Linear	Not assessed	No	
de Menezes et al	2010	Linear	Cheilion	Yes	
Fourie et al	2011	Linear	Cheilion, alare, tragus, pronasale, subnasale	Yes	Stereophotogrammetry was worse than CT
Metzger et al	2013	Linear	Exocanthion, gonion, labiale superior & inferior	No	
Othman et al	2013	Linear	Not known	No	
Kook et al	2014	Linear	Exocanthion, cheilion	Yes	
Chen et al	2015	Linear	Upper third of face	Yes	
Dindaroglu et al	2015	Linear, Angular	Cheilion, subnasale, tragus	No	ICC 0.85 to 0.99 for angular measurements

***Table 6 Studies of accuracy and precision of stereophotogrammetry with further details and notes***

***\*Error is measured for linear measurements between points, co-ordinates of points, or angles between three or more points***

***^If yes, studies were conducted with landmarks marked on the faces directly for use in all methods of face shape measurement***

### 3.1.4 VALIDATION STUDIES OF CANFIELD VECTRA CR10 CAMERA

The Canfield Vectra CR or Vectra CR10 stereophotogrammetry camera was used by me and this specific camera has been studied by a number of other groups for accuracy and reproducibility.

One group, based at the University of Milan, had previous experience of researching accuracy of stereophotogrammetry using older cameras (de Menezes et al., 2009). Using a cubic object, they found a mean absolute error of 0.02mm (0.2%) and of 0.04 degrees (0.04%) for linear and angular measurements respectively between the Canfield Vectra CR10 and direct measurement (de Menezes et al., 2010). They then assessed reproducibility in ten young adult human faces for 50 landmarks, including 15 out of the 22 landmarks used in my study: exocanthion (2), endocanthion (2), nasion, pronasale, subnasale, alare (2), labiale superius, labiale inferius, cristae philtri (2), cheilion (2). They also use three further landmarks close to landmarks in my study: menton for gnathion and subaurale for lower auricular attachment (2). Mean reproducibility error was 0.29mm - 0.53mm. Cheilion-cheilion distance was the least reproducible in all analyses but the mean reproducibility error was always less than 1.2mm. Note that the authors had marked all landmarks on the subjects' faces before image capture, so landmark placement errors were not part of the reproducibility assessment.

The same group later showed that the same camera could be used to assess the angle of the occlusal plane (the plane of intersection of the upper and lower teeth) using a subset of the same facial landmarks, with a mean interoperator error of 0.9-1.9 degrees (Rosati et al., 2012). The group then used 3D face surfaces and superimposed a 3D dental cast image from laser scanning. They found that the mean absolute error for distances between facial landmarks and dental landmarks was less than 1.7mm compared to *in vivo* measurements. They have also used the Canfield Vectra camera to

look at lip protrusion and facial aesthetics using facial landmarks but without any further assessment of accuracy or reproducibility (Rosati et al., 2014; Rossetti et al., 2013).

Another group compared the accuracy of stereophotogrammetry with direct anthropometry and 2D photographs. They do not specify the stereophotogrammetry camera model that they use but from the figures in their manuscript, it appears to be a Canfield Vectra CR10 system (Ghoddousi et al., 2007). They found a mean absolute error of 0.23mm for facial landmark distances using stereophotogrammetry compared to direct anthropometry. Others have used similar Canfield Vectra CR10 cameras to measure breast shape, except with three camera pods instead of two (Quan et al., 2011; Roostaeian and Adams, 2014). They did not perform an assessment of the camera's accuracy.

These studies show that accuracy and reproducibility in the Canfield Vectra system is unlikely to be significantly different to the other stereophotogrammetry cameras. Furthermore, they are reassuring given that one group used very similar facial landmarks to my own work. Even so, it is not clear exactly how their camera was set up. Background lighting, subject positioning and subject expression are likely to differ from my study. Landmarks were marked on the skin beforehand, which I did not do here. Also my subject population included a much broader age range and people with illness. Therefore, I have conducted a study of image reproducibility as part of this thesis to address these issues.

### 3.1.5 IMAGE CAPTURE AND REGISTRATION ERROR

Three studies specifically looked at the reproducibility of 3D image capture of the face at two or more prolonged time intervals. The first two studies used laser scanning and structured light systems and found accuracy within 1mm for time intervals between 3 days and 3 weeks – although one study used plaster casts (Kau et al., 2005; Ma et al.,

2009). The third study used 3D stereophotogrammetry at intervals of 1 minute and 3 weeks in 15 adult subjects, without using landmarks. The authors found that the mean system error in image capture and registration was less than 1.1mm looking at over 20,000 point co-ordinates on the face surface (Maal et al., 2010). They then found that in repeated images of one subject's face over 6 weeks, the mean error between all point co-ordinates was 0.36mm (Maal et al., 2011).

This suggests that repeated image capture in itself does not introduce significant errors, even at time intervals of up to three weeks. This is important for two reasons: in my work, I have used stereophotogrammetry to recruit subjects over two years and in one particular study I explore the change in face shape with age longitudinally.

### 3.1.6 HEAD POSITION

Another source of error during image capture is the change in a subject's position. One published study of head position during clinical photography found that adult head position was reproducible even without a mirror (Chiu and Clark, 1991). With stereophotogrammetry, the 3D image can be manipulated on screen into a desired position so head position should be irrelevant. One caveat is that the subject should face the field of view of the cameras. Looking off centre may lead to areas of the face missing in the 3D image. Two studies of head position in stereophotogrammetry using a 2-camera system suggest that accuracy is not affected if horizontal rotation of the face is less than 15-30 degrees from the midline (Lee et al., 2004; Lübbers et al., 2010).

However, both studies used mannequins and there has been no assessment of the effect of head position and orientation for human face images. Therefore, I have included head position in my own study of reproducibility. It is especially relevant to my subject population which includes children and people with intellectual disability, both of whom may not be able to maintain the ideal head position during image capture.

### 3.1.7 FACIAL EXPRESSIONS

Five studies have assessed the effect of different facial expressions on stereophotogrammetry. Using facial landmarks, two studies showed that a neutral facial expression was the most reproducible with a mean error less than 0.74mm (Johnston et al., 2003; Sawyer et al., 2009). Both studies had compared a neutral expression to smiling, cheek puffing, lip pursing and eye closure, using 15 identical or very similar landmarks to my own study.

Maal and colleagues used a different method to landmarking in which they analysed the root mean square difference for each of over 20,000 point co-ordinates in different facial expressions. They found that a laughing facial expression tripled the mean error in all point co-ordinates to 1.5mm compared to a neutral expression (Maal et al., 2010). In their subsequent study of one subject over 6 weeks, they looked to see if errors from facial expression variation could be reduced by asking the subject to wear a wax bite during both instances of image capture. The mean error did not vary with and without the wax bite (0.35mm vs 0.36mm) and error was maximal around the eyes and the mouth (Maal et al., 2011). The final study aimed to look at the effect of involuntary facial movements. In an unusual protocol, two of the study authors underwent repeated stereophotogrammetry to show the effect of 'involuntary' movements between photos on landmark positions (Lübbbers et al., 2012). They conclude that involuntary movements increase error up to 3.3mm for some measurements compared to a mannequin. However, their conclusion seems unwarranted given that the subjects stated they adopted a neutral pose just before image capture. There are also likely to be other differences between a human face and a mannequin head aside from facial expressions.

Taken together, the research suggests that there is good reproducibility of a neutral facial expression in stereophotogrammetry, and even maximal facial distortion from

laughing or cheek puffing causes small errors of approximately 1-2mm. The studies do not assess how much change there is in landmark positions due to facial expressions though. Therefore, I have also assessed the reproducibility of measurements using landmarks and using FSD with different facial expressions.

### 3.1.8 SUBJECTS WITH FACIAL DYSMORPHISM

Three of the validation studies used subjects with dysmorphic facial features, which can affect the accuracy of stereophotogrammetry. Facial shape may not be captured fully, such as in 3D images that omit low set ears. Distortion of facial features can make it more difficult to place landmarks, such as ptosis obscuring the view of the outer canthus and exocanthion.

The three studies each used a specific population of people with dysmorphism: infants with cleft lip, children with Down syndrome or craniosynostosis, and children with 22q11.2 deletion syndrome (Aldridge et al., 2005; Ayoub et al., 2003; Heike et al., 2009). Two of the studies do not comment upon the effect of dysmorphism on accuracy and precision, but Heike and colleagues state that they found no difference between subjects with and without 22q11.2 deletion syndrome in terms of accuracy. However, their study included 20 subjects with 22q11.2 deletion with a mean age of 11 years and 20 control subjects with a mean age of 25 years. Therefore, comparable control subjects had not been used and the sample was small overall. They point out that the syndrome causes subtle facial changes and that accuracy may be affected significantly in other conditions because of landmarking error (Heike et al., 2009). I study a population of people with epilepsy in whom some subjects were known to have facial dysmorphism. These previous studies suggest that dysmorphism itself should not affect the accuracy of landmark placement.

### 3.1.9 LANDMARKING ERROR

Landmark placement or 'landmarking' is the major operator-dependent step in stereophotogrammetry and DSM creation in my studies. Thus it was very important to look at errors in this step.

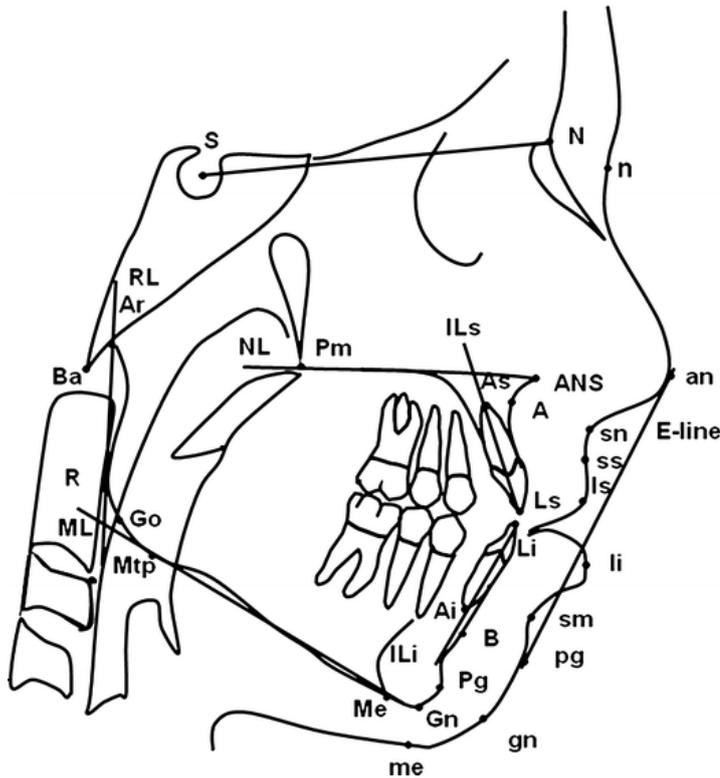
Twelve of the stereophotogrammetry studies described above relied on palpation and identification of landmarks on a subject's face directly, much as in direct anthropometry. When using stereophotogrammetry, they did not have to find the position of landmarks on the 3D images as the faces were already marked. Therefore, they did not assess for errors due to landmarking. The remaining studies placed landmarks on the face images. Five of these studies measured linear distances between these landmarks. Unfortunately, this means one cannot estimate the error in the placement of each landmark itself. For example, if two landmarks have been placed 5mm superiorly to their true locations, the distance between them would be the same as if they had been placed accurately. Three studies assessed point co-ordinates and specifically explored landmarking errors.

One study compared landmark placement on facial casts between a direct co-ordinate measuring machine and stereophotogrammetry (Khambay et al., 2008). Co-ordinate measuring machines (CMMs) use a moving probe to specifically measure co-ordinates in space, in this case with a manufacturer's reported accuracy of 0.0005mm, but they are impractical to use on a human face. The authors found that the CMM reproducibility was 0.1mm and using the CMM as the gold standard, the mean accuracy of stereophotogrammetry landmarking was 0.19-0.23mm. Intra-operator reproducibility error when placing landmarks was 0.06-0.07mm. Another study has found that 75-92% of facial landmarks (11 of 24 were also used in my study) have an intra-operator error of <1mm, which falls to 25-46% of landmarks for inter-operator error of <1mm (Gwilliam et al., 2006). A third group found intra-operator and inter-operator

reproducibility to be similar at 0.65-1.00mm and, of their 49 landmarks, fifteen are also used in my study (Plooij et al., 2009).

Landmark placement appears to be accurate for a subset of the landmarks used in this study, using different cameras. However, the traditional descriptions of landmark locations do not allow accurate placement because they are designed for direct anthropometry. Hence it is possible that the same landmark has been placed in slightly different locations in different studies depending on the features used. Such variation and interpretation is another reason to conduct my own study of reproducibility. I have fully described the features used to place landmarks in my study in Chapter 2.

One other study to note is of landmark reproducibility in 3D face images obtained from laser scanning (Toma et al., 2009). The authors assessed landmark reproducibility for 19 of the 22 landmarks used within this study. They found that intra-operator reproducibility was <1mm for 89% of co-ordinates and inter-operator reproducibility was <1mm for 83% of co-ordinates. They used their findings to rank all co-ordinates in order of reproducibility and the worst five were exocanthion (left and right), palpebrale superius (left and right), glabella, nasion and pogonion. Of these, glabella and pogonion were not used here, but gnathion is in a similar location to pogonion (Figure 21).



**Figure 21** A cross-section of the face showing bony and soft tissue landmarks. The relevant landmarks here are: Gn = bony gnathion, gn = soft tissue gnathion (used in my work), Pg = bony pogonion, pg = soft tissue pogonion. Soft tissue pogonion and soft tissue gnathion are both landmarks of the chin in the midline, and gnathion is likely to show similar reproducibility in placement to pogonion. Image used with written permission from Springer (Bartzela et al., 2012)

Nasion and exocanthion have been found to be the least accurate or precise in some stereophotogrammetry studies too (Table 6). Given these findings, extra care has been taken to use clear definitions for these landmarks and also to optimally view images (on a large monitor with magnification and rotation used as needed) before landmark placement. I also reassessed landmark reproducibility and particularly the effect of landmarking error in dense surface models and especially FSD, which has never been studied before to my knowledge.

### 3.1.10 DENSE SURFACE MODELLING ERROR

Dense surface models (DSMs) have been used in face shape analysis for over a decade (Hammond et al., 2004). However, no stereophotogrammetry study has assessed the effect of errors in head position, image capture or facial expression on the generation of

DSMs. One study has looked at the effect of landmarking error on dense surface modelling but used different landmarks to my studies (Fagertun et al., 2014). In this study, the authors selected 36 adult face images taken using a Canfield Vectra M3 system, which uses three camera pods. The 3D face images were initially converted to 2D images for automatic detection and annotation of surface landmarks by specialised software. They were then transformed back to 3D images and experienced operators manually checked and adjusted landmarks. Another difference to the current study is that the authors used 73 landmarks, only 24 of which are classical landmarks and 49 were termed “pseudo-landmarks”. Overall, there was a mean intra-operator reproducibility error of 1.5mm. Next, they chose different subsets of the 73 landmarks and created dense surface models. They analysed which of the DSMs predicted the location of the remaining landmarks better than the operators. Intriguingly, the mean error variance for every point on the DSM was reduced to 0.54mm when the model was made from only 14 particular landmarks instead of the 73-landmark DSM. Also, the 14-landmark DSM could then predict the position of 16 other landmarks more precisely than the operators. There are three caveats to these findings though. The authors do not mention, although it is evident in their data, that their 14-landmark DSM also predicts over 50 landmarks with less precision than the 73-landmark DSM. Secondly, much of the benefit of their 14-landmark DSM was for increasing the accuracy of the ‘pseudo-landmarks’ on the jaw, which are not used in most other studies, including this current one. Thirdly, a reduced landmark set may not be appropriate for study populations including people with facial dysmorphism or children.

### 3.1.11 SUMMARY

Therefore, given the different errors that are possible with stereophotogrammetry, which have not been fully described for my patient population, my camera system and my landmarks, I have conducted a study of intra-operator reproducibility. Although all analyses were performed by me, I also assessed inter-operator reproducibility between

me and another operator (Peter Hammond). The purpose of inter-operator reproducibility was to consider future applications of this technology in which multiple operators may be landmarking images.

I then assessed particular sources of error in more detail. Firstly, I looked at image capture and registration error. Then I explored the effect of head position systematically in three dimensions. Lastly, I investigated the effect of facial expressions. For these detailed assessments, the effect on dense surface models was also investigated.

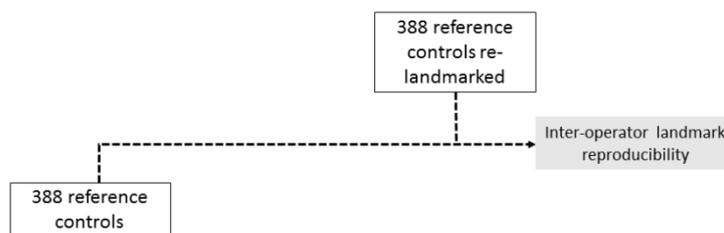
## 3.2 HYPOTHESES

I hypothesise that landmark reproducibility using the camera and associated software is similar to that seen in previous studies. Furthermore, I hypothesise that FSD, the primary endpoint in the following studies, is robust to variables in image capture, such as calibration, orientation and facial expression.

## 3.3 METHODS

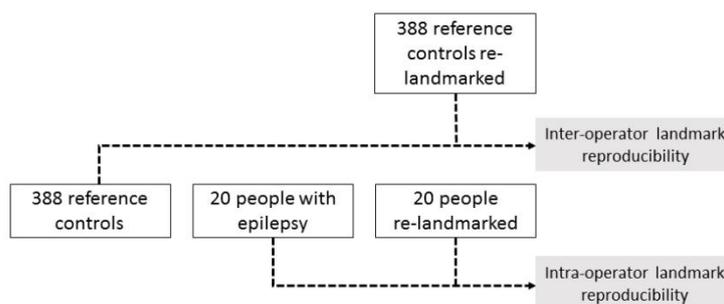
### 3.3.1 SUBJECTS AND DENSE SURFACE MODELS

For this study, three different sets of subjects were used and two different DSMs were created. An overview of the groups is shown later in Figure 26. Set 1 was of control subjects before I started any data collection from patients (Figure 22). Their images had been captured during a previous study (Hammond et al., 2005). Twenty images were chosen randomly, using an online random number generator (available at <http://www.random.org>), and used to assess inter-operator reproducibility of landmark placement for two operators (Peter Hammond and Vamsi Chinthapalli). The purpose of this analysis was to identify any significant errors in reproducibility before data collection proceeded.



**Figure 22** *Inter-operator reproducibility was assessed in the group of reference subjects between two operators. A random selection of 20 face surface images, called Set 1, from the total of 388 images was chosen with landmark placements compared.*

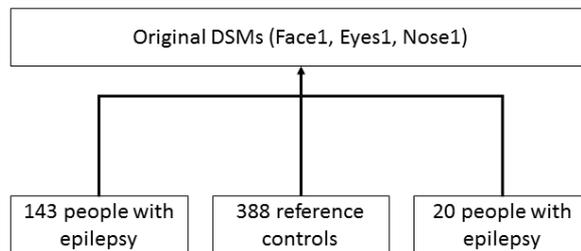
Set 2 comprised twenty adults with epilepsy recruited in my study (Figure 23). They were selected randomly from Group 1 using the aforementioned online random number generator (available at <http://www.random.org>), the first group recruited for my study, which comprised 163 people with epilepsy. As noted earlier, few studies have looked at reproducibility of landmarks in people with facial dysmorphism and none have looked at reproducibility of DSMs. For this set, I assessed intra-operator reproducibility of landmark placement.



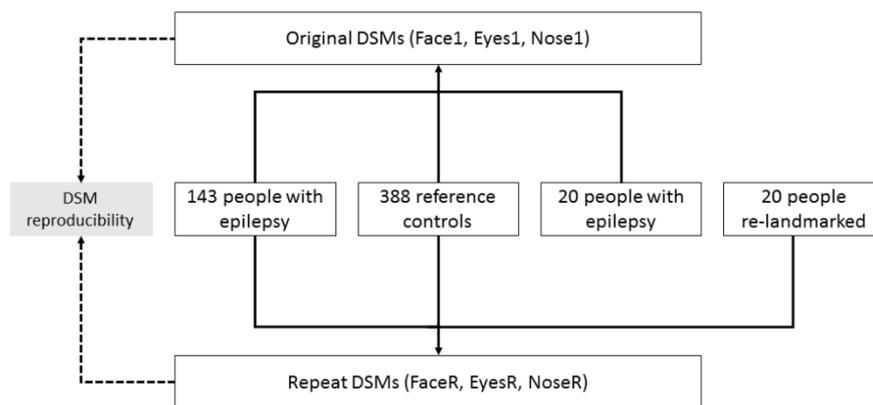
**Figure 23** *Intra-operator reproducibility was assessed by looking at a random selection of 20 people with epilepsy, Set 2, within the total sample of 163 people with epilepsy in Group 1. The same operator placed landmarks twice with an interval of six months to minimise recall bias.*

Using the same set of adults who had repeat landmarks applied, I explored the effect of landmark placement on FSD. For each region of the face, two DSMs were created using

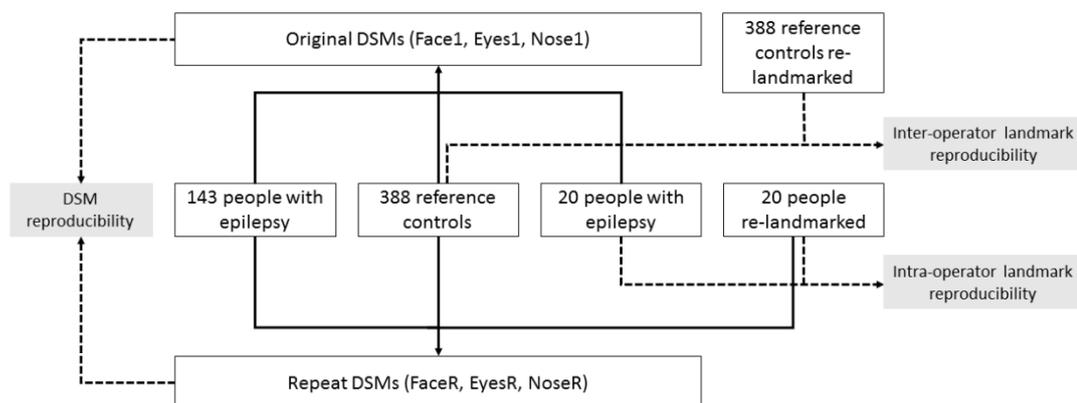
the full study cohort at the time. Three regions of the face were assessed and the original DSMs with the original landmark placements were called Face1, Eyes1 and Nose1 (Figure 24). The repeat DSMs with the revised landmark placements were called FaceR, EyesR and NoseR – ‘R’ signifying ‘Reproducibility’ DSM. The only difference between Face1 and FaceR was the location of landmarks in the face images of the chosen twenty subjects; as was the case for Eyes1 and EyesR, and Nose1 and NoseR (Figure 25). Therefore, any effect on FSD in FaceR will be due to variation in landmark placements; likewise for EyesR and NoseR.



***Figure 24 Three DSMs (Face1 for whole face, Eyes1 for periorbital region, Nose1 for perinasal region) created using 163 people with epilepsy (Group 1) and 388 reference subjects. Within this group, 20 people with epilepsy, Set 2, had been randomly chosen to be part the intra-operator reproducibility error study, but not the remaining 143 people with epilepsy.***



**Figure 25** In the FSD reproducibility study, the previous set (Set 2) of 20 adults with epilepsy and repeat landmark placements were now included in new DSMs of the same regions – FaceR, EyesR and NoseR. The only difference between Face1 DSM and FaceR DSM is that 20 people had their landmarks replaced in FaceR. Everything else was kept constant, so this study only looks at how landmark reproducibility affects DSM creation and subsequent calculation of FSD.



**Figure 26** Overview of reproducibility studies looking at landmark placement and intra-operator error, inter-operator error and arguably most importantly, the DSM reproducibility study.

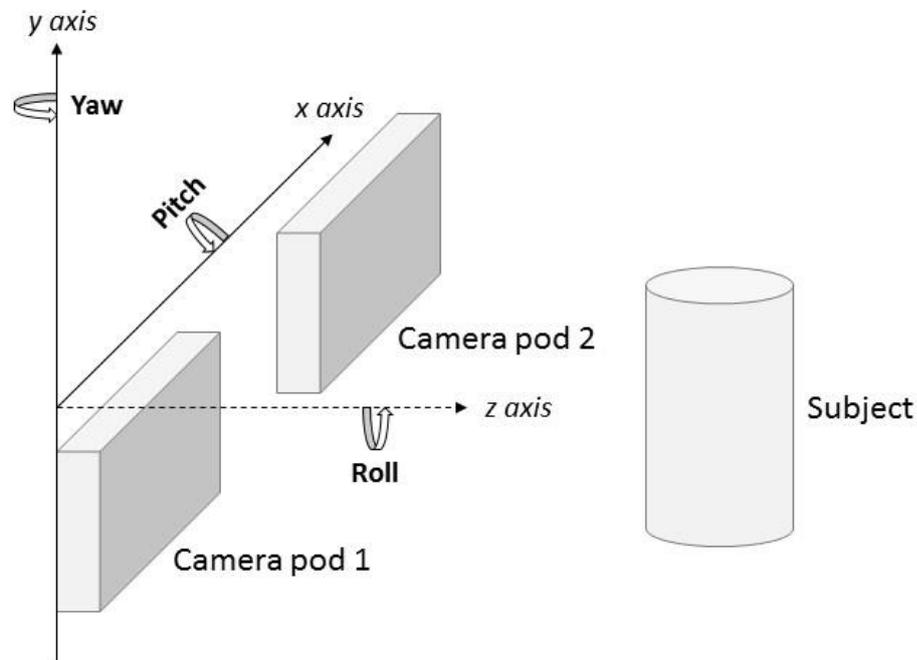
A third set, Set 3, of twenty adult subjects was then used to investigate the effect of calibration, head positioning and facial expressions. Eight of these people were also people with epilepsy who had been recruited into my later studies. The third set was incorporated into a separate set of DSMs (Face3, Eyes3, Nose3, Mouth3) along with the Group 3 study cohort of people with and without epilepsy. The effect on FSD was

assessed due to repeat calibration, variation in head position and variation in facial expressions.

### 3.3.2 IMAGE CAPTURE AND LANDMARKING

In the first set of control subject images, three different stereophotogrammetry systems had been used for image capture including 3dMD DSP400, Genex Facecam and Canfield Vectra CR10. In the second and third sets, subjects were prospectively acquired and images were all taken using one Canfield Vectra CR10 system. For the second set, only one image was taken per subject using the methods described in Chapter 2.

For the third set only, image acquisition comprised eleven images per subject (Table 7). All eleven images were taken in the same photography session over approximately 30 minutes. As far as possible, lighting and position was kept as similar as possible between images. Image 1 was in a neutral position and neutral facial expression. Images 2-7 were taken at six different angles, two per axis of rotation. For these, the camera system was rotated because this was more practical and measurable than attempting to alter the subject's head position itself. The camera was mounted on a tripod using a Manfrotto 410 adjustable junior gear head with spirit level and angle markings in one degree increments (Cassola, Italy: Lino Manfrotto + Co. Spa). This tripod mount could be rotated in three dimensions and the axes of rotation are shown in Figure 27. The horizontal x-axis runs between both camera pods and the vertical y-axis runs between individual cameras within a pod. The z-axis is perpendicular to the other two axes.



**Figure 27** Axes of rotation of the camera used to assess head position relative to the camera using aviation terminology. Yaw is rotation around a vertical column, in this case the centre of the tripod base. Pitch is rotation around a horizontal line between the two pods. Roll is rotation in a plane perpendicular to both of the other rotations described and is around an imaginary line between the subject and the centre of the camera.

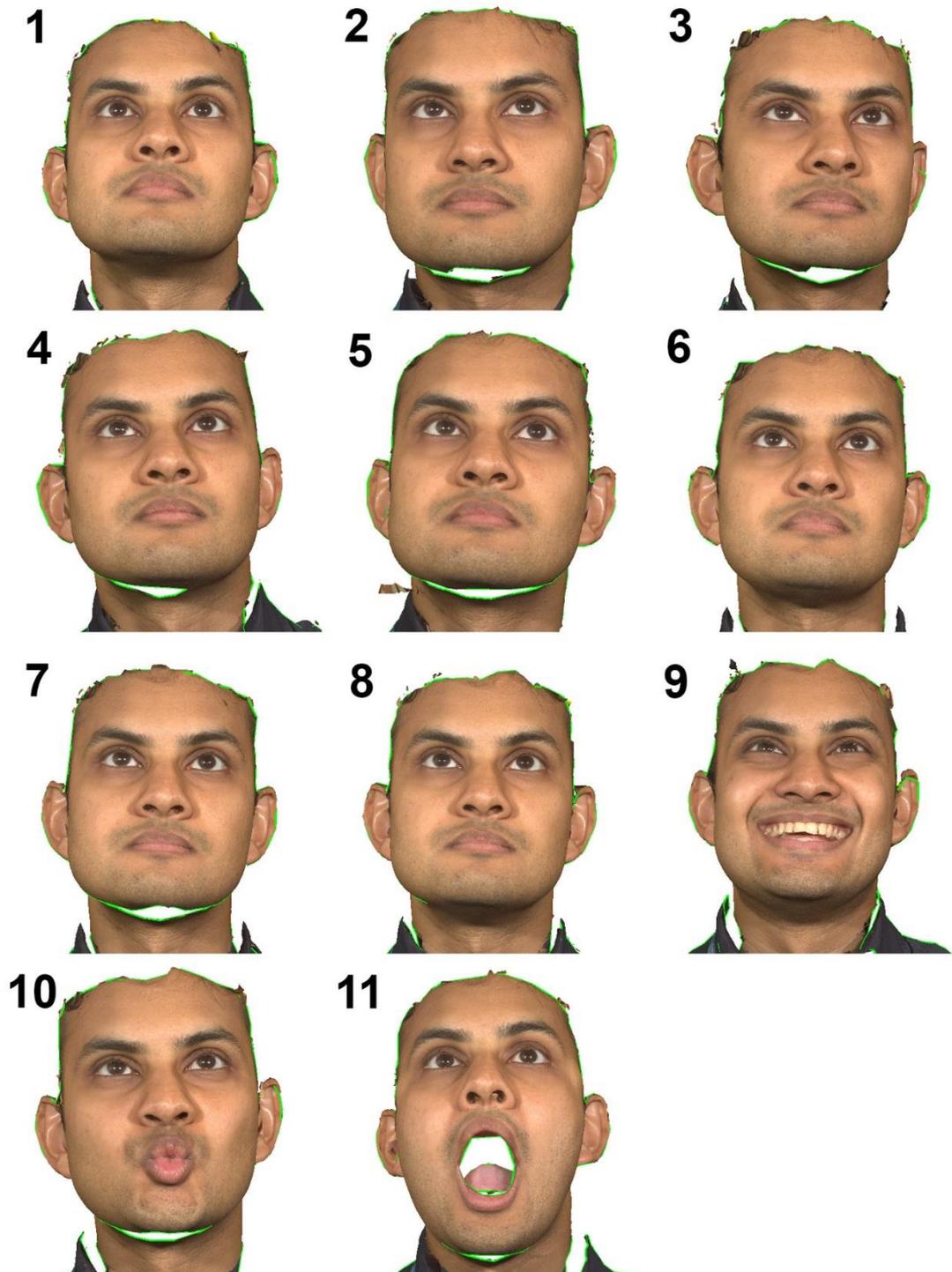
Rotation can be described using aviation terminology as pitch (along x-axis), yaw (along y-axis) and roll (along z-axis). I chose an angle of 5.0 degrees clockwise and 5.0 degrees anticlockwise for each axis as the size of rotation. This angle was measurable using our tripod and corresponded to visually noticeable rotation of the face. For pitch, the camera was orientated to have a baseline pitch of -20 degrees clockwise, which is well-recognised to aid coverage of the subnasal region of the face (Ghoddousi et al., 2007; Heike et al., 2010). All changes in pitch described below are in relation to this baseline.

After these images, Image 8 was taken back in the neutral position after re-calibration of the camera system. Images 9-11 were taken with different facial expressions. These were taken to see the maximal effect that expressions could have on face shape. Therefore, subjects were asked to open their mouth fully (Image 11), similar to the laughing posture previously found to cause the largest error in registration (Maal et al., 2010). Smiling fully with lips open (Image 9) was chosen as a reproducible and natural facial expression, which has also been studied before (Johnston et al., 2003; Sawyer et al., 2009). Finally lip puckering was chosen as a different and extreme example of facial expression (Image 10). These images were compared to Image 8, of the face in a neutral expression. Image 1 was not used as the comparison because re-calibration (performed prior to Image 8) could affect the results.

The sequence for these images is shown in Table 7 and the images themselves are shown for one participant (Vamsi Chinthapalli) in Figure 28.

Image	Pitch (°)	Roll (°)	Yaw (°)	Expression	Calibration
1	0	0	0	Neutral	Yes
2	0	0	-5	Neutral	No
3	0	0	5	Neutral	No
4	0	-5	0	Neutral	No
5	0	5	0	Neutral	No
6	-5	0	0	Neutral	No
7	5	0	0	Neutral	No
8	0	0	0	Neutral	Yes
9	0	0	0	Smile	No
10	0	0	0	Pucker	No
11	0	0	0	Open mouth	No

***Table 7 Sequences of images taken per participant for the third set. Head position was assessed relative to the camera in terms of camera pitch, roll and yaw in degrees clockwise. Then images were taken after re-calibration in different facial expressions: neutral, smile with lips open, lips puckering and mouth open.***



**Figure 28** Raw 3D face images for one participant in the third set to illustrate the sequence described in the methods. Images 2-7 show the effect of camera rotation. Image 8 is the same as image 1 but with re-calibration. Images 9-11 show different facial expressions: smiling with lips open, lip puckering and mouth open.

A final detailed assessment of head position was performed in one subject (Vamsi Chinthapalli). For this, the subject underwent capture of 125 3D face images in one

photography sitting over six hours. This consisted of images taken at degree intervals of -10, -5, 0, 5, 10 for each of the three axes x, y and z (pitch, yaw, roll) in every combination. To aid accuracy, the subject maintained the same face position relative to the camera by resting one point on his chin on a second tripod mount, which was fixed. This location did not interfere with any landmark placements. In addition, the tripod was automatically excluded from the face models used in DSM creation because it was outside the lower boundary of the face used in DSM creation, which is typically close to gnathion. These images were also included with the twenty control subjects in four DSMs (Face3, Eyes3, Nose3, Mouth3) to look for differences in FSD.

Landmarking was performed using all 22 landmarks previously described. All DSMs were created using the methods described in Chapter 2. Landmark abbreviations are used as in Chapter 2: exocanthion (exo L/R), endocanthion (endo L/R), palpebrale superius (psup L/R), palpebrale inferius (pinf L/R), nasion (n), pronasale (prn), subnasale (sn), alari nasi (ala L/R), labiale superius (ls), labiale inferius (li), cristae philtri (chp L/R), cheilion (ch L/R), gnathion (gn) and otobasion inferius (oto L/R).

### 3.3.3 INTER-OPERATOR REPRODUCIBILITY ERROR

Inter-operator reproducibility was assessed by calculating the Euclidean distance between the two placements of for each landmark. When landmarking in Facemark, Cartesian co-ordinates are automatically created in three dimensions and the Pythagorean formula may be used:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

d = Euclidean distance between first and second placement of landmark

(x<sub>1</sub>,y<sub>1</sub>,z<sub>1</sub>) = first operator placement of landmark

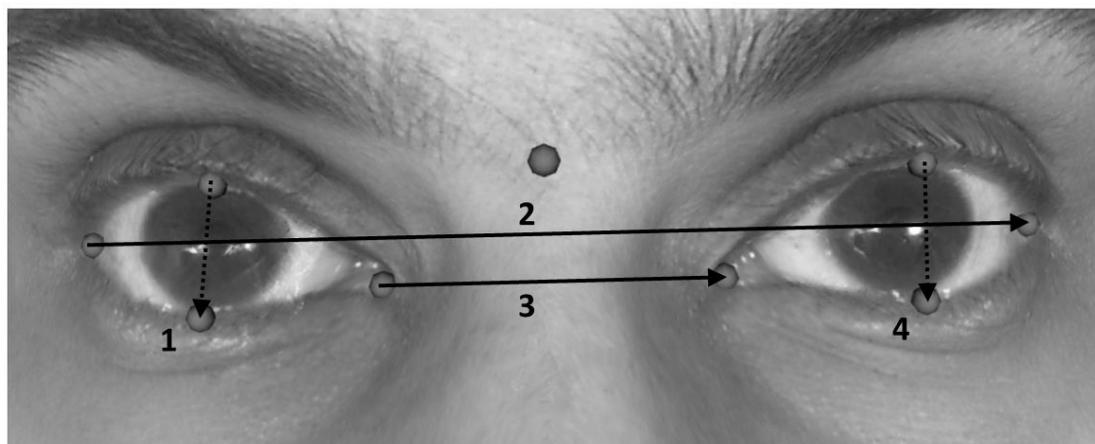
(x<sub>2</sub>,y<sub>2</sub>,z<sub>2</sub>) = second operator placement of landmark

For each landmark, this distance was measured for every face image and then the mean and standard deviation were calculated.

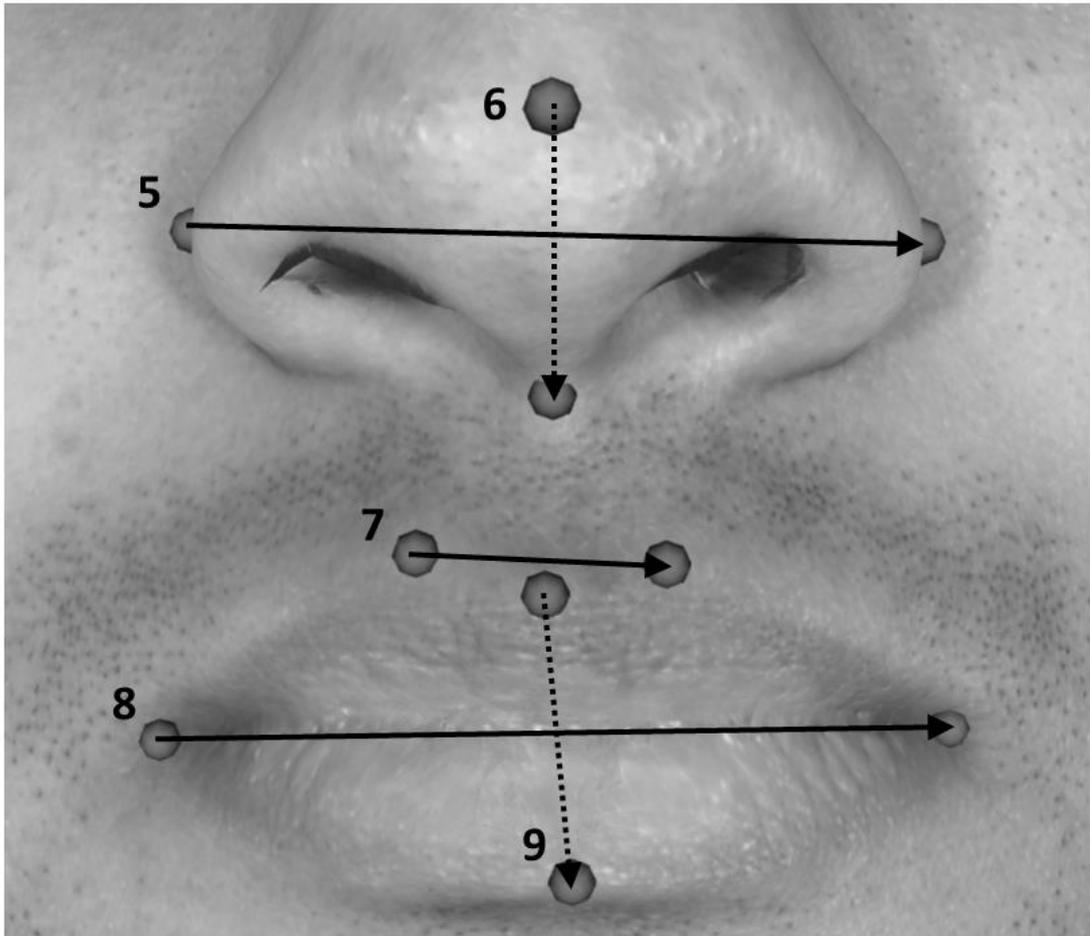
In addition, 12 inter-landmark distances were measured using the same formula above between pairs of landmarks. The distances used are in Table 8 and are also seen in Figure 29, Figure 30 and Figure 31. All have been used in previous studies except for palpebral fissure height and face width (de Menezes et al., 2010; Ghoddousi et al., 2007; Heike et al., 2009; Kook et al., 2014; Metzger et al., 2013; Weinberg et al., 2006; Wong et al., 2008). Palpebral fissure height uses two landmarks, palpebrale superius and palpebrale inferius, not used in other stereophotogrammetry studies of reproducibility, but used in other studies of facial dysmorphism (Hammond et al., 2005; Tassabehji et al., 2005). Orobasion inferius is also not used in other reproducibility studies but many use a nearby landmark, either tragion or subaurale, to measure face width. The reason for using orobasion inferius in this study is explained in Chapter 2. The distances use every landmark at least once. For inter-landmark distances, the 2-way mixed models intraclass correlation coefficient (ICC) was calculated.

Inter-landmark distances	Name
R exocanthion – L exocanthion	Biocular width
R endocanthion – L endocanthion	Intercanthal width
R palpebrale superius – palpebrale inferius	R palpebral fissure height
L palpebrale superius – palpebrale inferius	L palpebral fissure height
R alare nasi – L alare nasi	Nostrils width
R crista philtrum – L crista philtrum	Philtrum width
R cheilion – L cheilion	Labial fissure width
Labiale superius – labiale inferius	Lip height
R otobasion inferius – L otobasion inferius	Face width
Nasion – gnathion	Face height
Subnasale – pronasale	Nose protrusion
Subnasale – gnathion	Lower face height

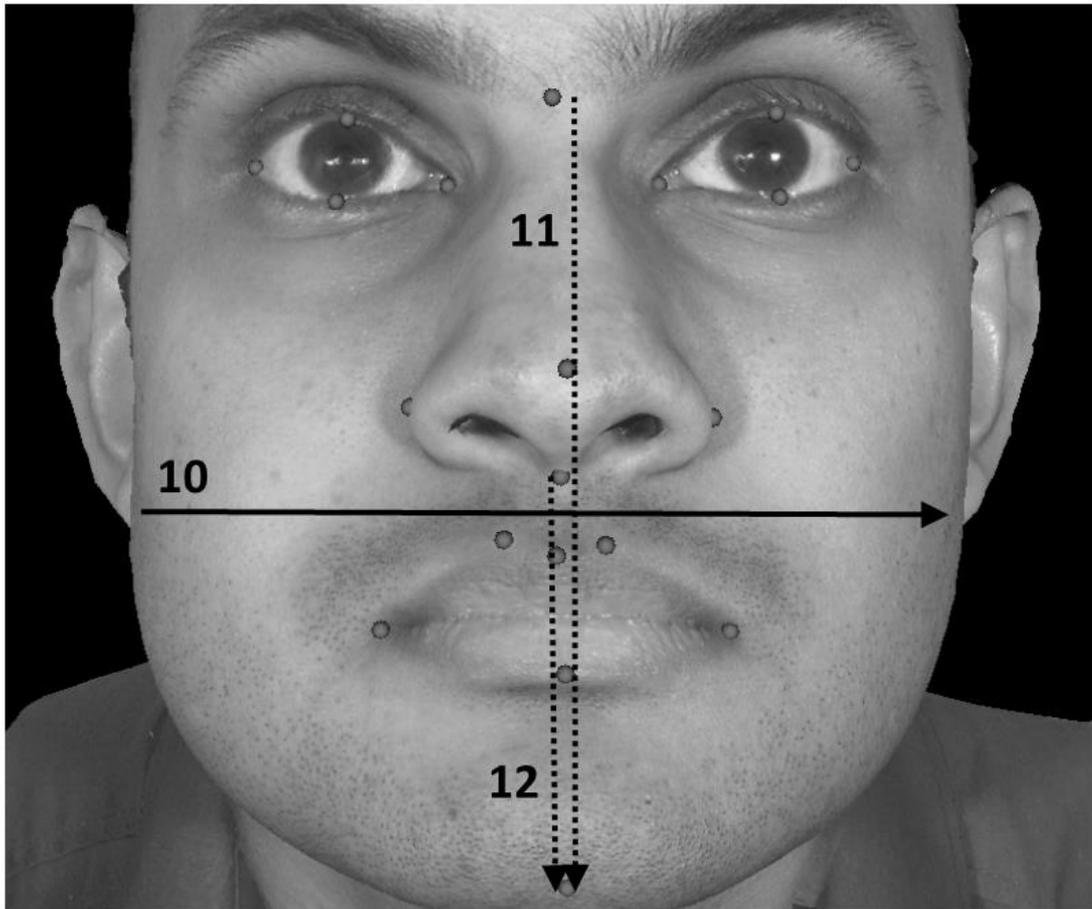
**Table 8** Inter-landmark distances measured for reproducibility error



**Figure 29** Four inter-landmark distances around the eyes are shown here. Each line is labelled and the start and end landmarks for each line are: 1 = p<sub>sup</sub>R to p<sub>inf</sub>R, 2 = exoR to exoL, 3 = endoR to endoL, 4 = p<sub>sup</sub>L to p<sub>inf</sub>L



**Figure 30** Five further inter-landmark distances of the nose and mouth regions. Each line is labelled and the start and end landmarks for each line are: 5 = alaR to alaL, 6 = prn to sn, 7 = chpR to chpL, 8 = chR to chL, 9 = ls to li



**Figure 31** Three more inter-landmark distances of facial dimensions used in the reproducibility study. Distances are: 10 = otoR to otoL, 11 = n to gn, 12 = sn to gn.

### 3.3.4 INTRA-OPERATOR REPRODUCIBILITY ERROR

This was assessed in a second set of subjects, all with epilepsy. One operator (Vamsi Chinthapalli) placed landmarks on the face images a second time, six months after the first placement. Reproducibility was then compared in the same way as for inter-operator reproducibility in the previous section, including using inter-landmark distances. For FSD values, the Shapiro-Wilks test of normality was performed to see if data were normally distributed. FSD values were not normally distributed in any of the six DSMs. Paired FSD values for each subject from the original and repeat DSMs were then tested to see if they were significantly different using a related samples Wilcoxon signed rank test for non-parametric data. In the three pairs of DSMs (Face1, FaceR, Eyes1, EyesR, Nose1, NoseR), the FSD values were also compared using 2-way Spearman's rank correlation coefficient ( $\rho$ ). This coefficient was chosen because the

data were not necessarily expected to have a linear relationship between the original and repeat DSMs.

In a separate analysis of FSD reproducibility, I compared FSD values between Image 1 and Image 8 in all of the subjects in Set 3. These images were part of the same DSMs, unlike the analysis in the paragraph above. This means the FSD values here should theoretically be identical for the same person. Hence, the differences in these FSD values for each subject show the reproducibility error in FSD due not just to landmarking, but to all the errors from repeat image acquisition. Therefore, it will include errors from changes in head position, face expression, camera calibration, image capture and registration, and repeat landmarking. Statistically, the median absolute difference (MAD) was used to describe this difference, where MAD is the median of the absolute difference between all FSD values and the median FSD value of the whole cohort. The FSD values between image 1 and image 8 were then also compared using a test of normality and then either a paired samples t-test or related samples Wilcoxon signed rank test.

### 3.3.5 CALIBRATION ERROR

Images in the third set of subjects were compared to assess if there was any additional error due to camera re-calibration. These are image 1 and image 8 in Table 7, which were of the same subject and taken in the same photography session. Two separate images are now being compared unlike in the first and second sets, in which the *same image* had different landmark placements by the same or a different operator. Using two images means that error in a landmark's absolute co-ordinates cannot be assessed because each image has a separate frame of reference and therefore different co-ordinates for the same point in space. Thus to compare the reproducibility, I only used inter-landmark distances.

### 3.3.6 HEAD POSITIONING

Errors due to head positioning were measured in the same way as for calibration error using inter-landmark distances. The four DSMs were then assessed (Face3, Eyes3, Nose3, Mouth3) and FSD values were not distributed normally. There are six face images per subject at different angles of rotation, two in each plane. All six were compared with each other and with the neutral position to assess for any significant difference in FSD using the related samples Friedman's two-way analysis of variance by ranks. This test is the non-parametric equivalent of the analysis of a two-way analysis of variance (ANOVA).

For the detailed analysis of one subject's images for five angles in each of the three axes, the independent samples Kruskal-Wallis test was used to assess if FSD varied significantly with camera rotation.

### 3.3.7 FACIAL EXPRESSIONS

Differences in inter-landmark distances due to facial expressions were calculated as for calibration error and compared to the distances when in a neutral expression. After a test for normality, an ANOVA test and post-hoc Tukey test were used to explore the relationship in any errors. The difference in FSD was assessed for each facial expression and for each of the four DSMs. A related samples Friedman's two-way analysis of variance by ranks was used for this.

## 3.4 RESULTS

### 3.4.1 SUBJECT CHARACTERISTICS

Each set had twenty subjects and age and sex characteristics are provided in Table 9. There were nine male subjects in set 1 and ten in each of sets 2 and 3. The age range was widest for set 1 (0.5 – 81.2 years) whereas in sets 2 and 3, all participants were adults except for one.

Subject	Set 1		Set 2		Set 3	
	Age	Sex	Age	Sex	Age	Sex
1	0.5	F	3.3	M	25.4	F
2	2.0	M	21.6	F	28.5	F
3	8.3	M	23.0	F	30.8	F
4	10.7	F	24.2	F	31.0	F
5	12.4	F	31.7	F	31.0	M
6	14.2	M	31.9	F	31.2	F
7	24.1	M	33.2	F	32.4	F
8	24.9	F	35.9	M	33.3	M
9	28.7	M	36.0	M	33.7	F
10	29.5	F	38.9	F	35.1	M
11	33.4	F	39.2	M	37.2	F
12	34.8	F	39.7	F	37.6	M
13	35.0	F	39.9	F	38.8	M
14	35.8	F	41.4	M	45.2	M
15	40.5	F	42.1	M	48.9	F
16	41.0	M	42.2	M	51.0	M
17	42.2	M	47.4	M	53.0	M
18	47.0	M	47.9	M	56.1	M
19	53.6	M	48.8	F	56.1	M
20	81.2	F	49.3	M	66.9	F

**Table 9** Age and sex characteristics of the twenty subjects analysed in each set

### 3.4.2 INTER-OPERATOR AND INTRA-OPERATOR REPRODUCIBILITY FOR LANDMARK POSITIONS

For inter-operator reproducibility in 20 control subject images from Set 1, seven landmarks were reproducible to within 1mm and twenty were reproducible to within 2mm (Table 10). For intra-operator reproducibility in 20 subject images from Set 2, I found that fourteen of the 22 landmarks were reproducible to within 1mm, and all were reproducible to within 2mm. Inter-operator landmarking showed a greater mean distance between landmark placements than for intra-operator landmarking for all except six landmarks: cheilion (left and right), otobasion inferius (left and right), nasion and gnathion. The alari nasi positions were particularly prone to variation in inter-operator comparisons.

Rank	Intra-operator			Inter-operator		
	Landmark	Mean (mm)	SD (mm)	Landmark	Mean (mm)	SD (mm)
1	ls	0.30	0.20	ls	0.58	0.32
2	chpL	0.49	0.40	chpL	0.63	0.29
3	exoL	0.55	0.38	endoL	0.83	0.50
4	sn	0.56	0.31	exoL	0.84	0.60
5	prn	0.62	0.41	otoR	0.86	0.51
6	pinfR	0.67	0.34	chL	0.96	0.85
7	psupR	0.69	0.72	exoR	0.99	1.18
8	endoR	0.71	0.49	pinfL	1.01	0.69
9	exoR	0.72	0.68	endoR	1.03	0.74
10	endoL	0.74	0.60	chpR	1.03	0.75
11	chpR	0.76	0.46	chR	1.03	0.80
12	li	0.76	0.56	pinfR	1.05	0.48
13	pinfL	0.80	0.54	otoL	1.09	0.71
14	psupL	0.97	0.55	prn	1.20	0.40
15	chL	1.05	1.18	li	1.06	0.60
16	chR	1.12	1.32	psupR	1.45	0.70
17	otoR	1.17	0.92	psupL	1.65	1.33
18	alaR	1.29	0.64	n	1.66	0.92
19	alaL	1.41	0.80	sn	1.67	1.15
20	otoL	1.72	1.51	gn	1.69	1.18
21	gn	1.77	2.00	alaR	2.86	1.04
22	n	1.84	2.92	alaL	3.57	1.12
	<i>All landmarks</i>	<i>0.94</i>	<i>0.82</i>	<i>All landmarks</i>	<i>1.31</i>	<i>0.77</i>

**Table 10** Intra-operator and inter-operator reproducibility in placement of the 22 soft tissue facial landmarks. The mean and standard deviation are shown for the discrepancy, in terms of Euclidean distance (mm), between the landmarks.

For inter-landmark distances, the overall mean inter-operator reproducibility error was 0.71mm and the overall mean intra-operator reproducibility error was 0.34mm.

The inter-operator reproducibility error was less than 1.5mm for any distance and was >1mm for three distances. Intra-operator reproducibility was within 1mm for all but

one of the distances. There was a high intraclass correlation coefficient for all measurements, with the lowest being for the philtrum width (distance between cristae philtri) at 0.94 (Table 11). The nose width measurement (right alare to left alare) had an intraclass correlation coefficient of 0.99-1.00. These results show good reproducibility of landmark placement, comparable to previous stereophotogrammetry studies.

Distance	Intra-operator			Inter-operator		
	Mean (mm)	SD (mm)	ICC	Mean (mm)	SD (mm)	ICC
exoR - exoL	0.31	0.97	0.99	0.37	1.00	1.00
endoR - endoL	0.06	1.19	0.97	1.20	1.24	0.98
psupR - pinfR	0.16	0.38	0.99	0.42	0.72	0.96
psupL - pinfL	0.16	0.41	0.99	0.60	0.66	0.97
alaR - alaL	0.22	0.31	1.00	0.62	0.79	0.99
chpR - chpL	0.37	1.19	0.94	0.86	1.22	0.94
chR - chL	0.94	2.28	0.95	0.13	1.87	0.99
ls - li	0.01	0.59	1.00	0.60	0.86	1.00
otoR - otoL	0.16	0.99	1.00	0.24	0.47	1.00
n - gn	0.54	3.12	0.99	1.00	1.50	1.00
sn - prn	0.02	0.74	0.99	1.44	1.47	0.97
sn - gn	1.18	1.87	0.99	1.04	1.78	0.98

**Table 11 Intra-operator and inter-operator reproducibility of inter-landmark distances with mean and standard deviation of the difference between both measurements and the intraclass correlation coefficient.**

### 3.4.3 INTRA-OPERATOR REPRODUCIBILITY IN FSD ACROSS TWO DSMs IN PEOPLE WITH EPILEPSY

I also assessed the effect of landmark placement on the subsequent dense surface model creation and FSD. Three pairs of DSMs were assessed for FSD reproducibility due to repeat landmarking using Set 2 subjects (Face1 to FaceR, Eyes1 to EyesR, Nose1 to NoseR). Landmarking is used in thin plate spline deformation of face surfaces for a dense correspondence creation. Therefore, it may affect the dense correspondence and subsequent principal component analysis. Different landmark placements in some subjects in a cohort would thus be expected to alter any given DSM and affect FSD for all images within the group. This is what was found as FSD values were not the same in

each pair of DSMs with the original and repeat landmark placements. Despite different absolute values, FSDs for each subject were not significantly different between DSMs of the face (Face1 vs FaceR;  $p=0.678$ ), the periorbital region (Eyes1 vs EyesR;  $p=0.496$ ) of the perinasal region (Nose1 vs NoseR;  $p=0.369$ ). There was also strong correlation between the FSD values in all three DSM comparisons (Table 12). In the DSMs of the whole face, the twenty images that had repeat landmark placements showed strong correlation between the old and new FSD values (Face1 vs FaceR,  $\rho=0.94$ ;  $p=1.48 \times 10^{-9}$ ). Correlation was similar for DSMs of the periorbital region (Eyes1 vs EyesR,  $\rho=0.94$ ;  $p=1.21 \times 10^{-9}$ ) and weaker for DSMs of the perinasal region (Nose1 vs NoseR,  $\rho=0.90$ ;  $p=1.02 \times 10^{-7}$ ). This suggests that FSD does not vary significantly when landmarks are repositioned and use different DSMs.

	Redo landmark cases (n = 20)	Significance	All subjects (n = 117)	Significance
Face1 & FaceR	0.94	$1.48 \times 10^{-9}$	0.99	$7.13 \times 10^{-114}$
Nose1 & NoseR	0.90	$1.02 \times 10^{-7}$	0.99	$6.74 \times 10^{-91}$
Eyes1 & EyesR	0.94	$1.21 \times 10^{-9}$	0.99	$1.04 \times 10^{-113}$

**Table 12 Spearman's rank correlation coefficients for correlation between FSD values in three different DSMs of the face, perinasal region and periorbital region. Correlation is shown for just the 20 face images that had repeat placement of landmarks in the revised DSMs and then for all subjects in the DSM.**

#### 3.4.4 OVERALL REPEAT IMAGE REPRODUCIBILITY OF FSD IN THE SAME DSM

Because two different DSMs are used for each face region in the previous section, the absolute FSD values could not be directly compared themselves. However, I was able to use another group of people to assess how much FSD varies in the same DSM due to errors in both repeat image acquisition and repeat landmarking of a subject.

For Set 3, I obtained 11 images for each of 20 control subjects that were all used in the same DSMs – i.e. Face3, Eyes3, Nose3 and Mouth3. In Set 3, image 1 and image 8 are taken of the same person in the same sitting at two different time intervals with camera

re-calibration in between. The only noticeable differences between the images should be changes in patient position, patient expression, camera re-calibration and landmarking. Comparing changes in FSD between image 1 and image 8 for all twenty subjects allows me to estimate the overall reproducibility error in FSD from repeated image acquisition of a subject. These twenty subjects had all of their images included in four DSMs: Face3, Eyes3, Nose3 and Mouth3. Within the four DSMs, data were not normally distributed for the FSD values of the 20 control subjects in the original images and/or the recalibrated images.

The median difference in FSD between repeat images for these subjects was: 0.10 within Face3, 0.02 for Eyes3, 0.06 for Nose3 and 0.10 for Mouth3. Because these values are meaningless without knowing the distribution of FSD in the whole model, this has also been calculated (Table 13). Table 13 shows that compared to the overall distribution of FSD in each of the four DSMs, the median reproducibility error from repeat image acquisition is up to 0.10 of FSD.

The FSD values were not significantly different between image 1 and image 8 in DSMs of the whole face (related samples Wilcoxon signed rank test,  $p=0.46$ ), periorbital region (related samples Wilcoxon signed rank test,  $p=0.68$ ), perinasal region (related samples Wilcoxon signed rank test,  $p=0.39$ ) and perioral region (related samples Wilcoxon signed rank test,  $p=0.94$ ). Therefore there is no significant reproducibility error in FSD between two separate images taken for each of twenty subjects.

	DSM			
	Face3	Eyes3	Nose3	Mouth3
Median	9.26	13.12	8.78	6.86
Minimum	6.22	8.04	5.84	4.10
1st quartile	8.35	11.50	7.79	5.97
3rd quartile	10.51	15.08	10.04	8.04
Maximum	31.33	42.72	38.53	21.72
MAD	1.02	1.80	1.07	0.97
<i>Median FSD error for repeat Set 3 images</i>	<i>0.10</i>	<i>0.02</i>	<i>0.06</i>	<i>0.10</i>

**Table 13 Median FSD reproducibility error in four DSMs is shown in the bottom row for repeat image acquisition and landmarking in 20 control subjects (Set 3). This error is contrasted with the overall distribution of FSD in the four DSMs for all people with epilepsy and relatives. The median FSD reproducibility error is approximately 1-10% of the overall median absolute difference (MAD) seen in FSD in different subjects. The DSMs are for the whole face (Face3), the periorbital region (Eyes3), the perinasal region (Nose3) and the perioral region (Mouth3).**

### 3.4.5 CAMERA CALIBRATION

Inter-landmark distances had a mean reproducibility error of between 0.05mm and 0.31mm before and after re-calibration of the camera system (Table 14). The errors are less than the mean intra-operator reproducibility for inter-landmark distances that was already assessed (Table 11).

Distance	Mean (mm)	SD (mm)	ICC
exoR - exoL	0.11	0.47	1.00
endoR - endoL	0.23	0.70	0.99
psupR - pinfR	0.25	0.63	0.88
psupL - pinfL	0.31	0.76	0.88
alaR - alaL	0.21	0.59	0.99
chpR - chpL	0.15	0.82	0.98
chR - chL	0.11	1.73	0.97
ls - li	0.05	1.32	0.98
otoR - otoL	0.31	1.35	1.00
n - gn	0.17	1.70	0.99
sn - prn	0.09	0.42	0.99
sn - gn	0.16	1.82	0.98

**Table 14 Reproducibility error of inter-landmark distances with mean and standard deviation of the difference between measurements taken before and after camera re-calibration**

### 3.4.6 HEAD POSITIONING ERROR

Head positioning may affect landmark or FSD reproducibility and has previously only been studied in mannequins. All twenty subjects in Set 3 were included in the study of head position relative to the camera. For each of the six different camera angles relative to head position, the overall mean reproducibility error in inter-landmark distance measurement was between 0.74mm and 0.85mm when compared to the distances measured in a neutral position (Table 15). Greater reproducibility error was seen in the longest distances: sn – gn, n – gn and otoR – otoL. No individual inter-landmark distance had a mean error greater than 2mm in any position. The least reproducible distance appeared to be palpebral fissure height. Intraclass correlation coefficients were calculated for each inter-landmark distance between the different camera angles and showed strong agreement throughout (ICC 0.96-1.00).

In all four DSMs, FSD values were not significantly different due to head position relative to the camera. Median FSD values are shown in Table 16. It would appear that whole face FSD is most affected by changes in camera roll and yaw, but not pitch. Also, the perinasal FSD seems to be least affected by any changes in head position, which is probably due to its central position allowing for detailed image capture at different camera angles.

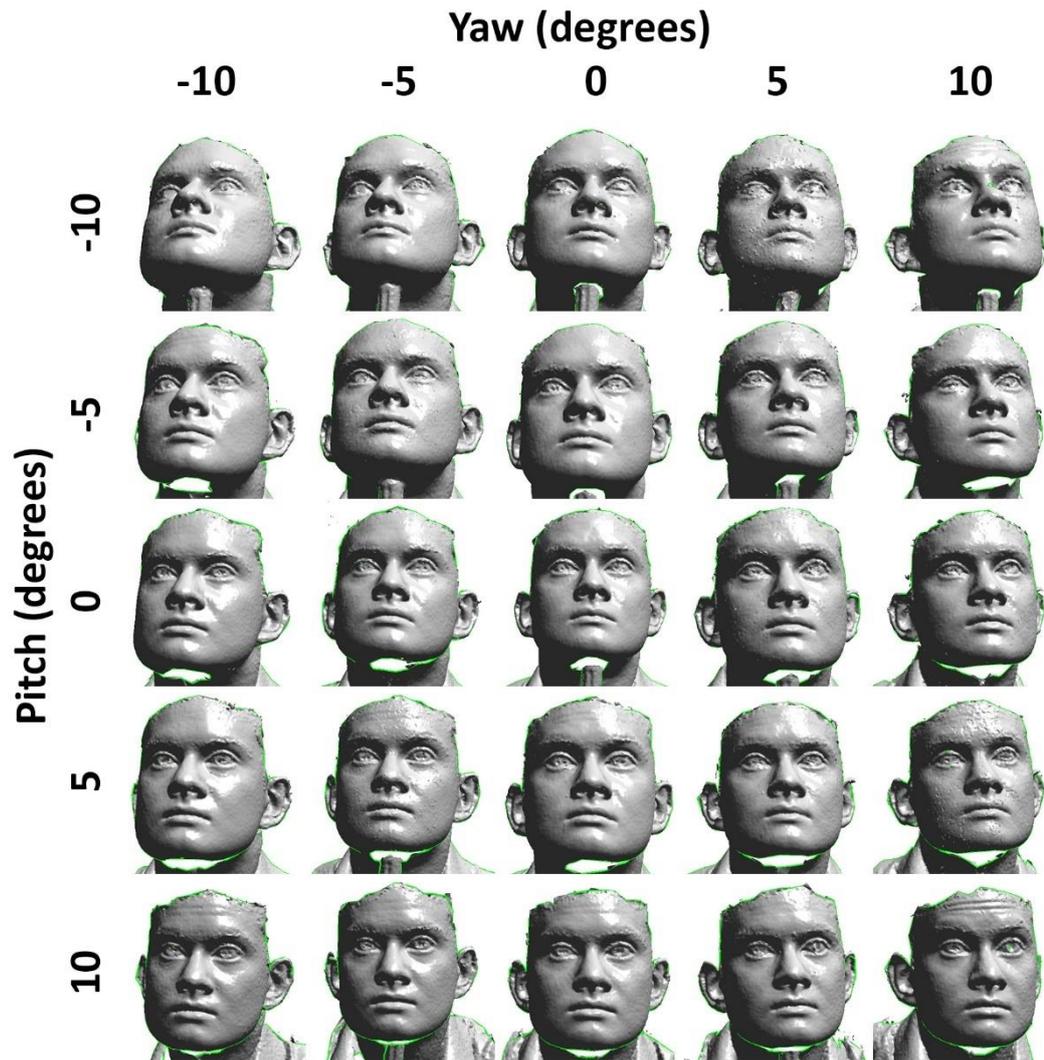
Distance	Pitch -5°		Pitch 5°		Roll -5°		Roll 5°		Yaw -5°		Yaw 5°		ICC
	Mean (mm)	SD (mm)											
exoR - exoL	0.37	0.48	0.46	0.57	0.40	0.50	0.57	0.69	0.29	0.37	0.50	0.68	<b>1.00</b>
endoR - endoL	0.64	0.81	0.63	0.87	0.52	0.79	0.51	0.75	0.42	0.50	0.59	0.74	<b>1.00</b>
psupR - pinfR	0.55	0.67	0.58	0.78	0.65	0.71	0.71	0.80	0.55	0.60	0.49	0.60	<b>0.96</b>
psupL - pinfL	0.59	0.80	0.61	0.71	0.69	0.77	0.74	0.87	0.67	0.76	0.67	0.74	<b>0.96</b>
alaR - alaL	0.41	0.53	0.52	0.60	0.54	0.64	0.57	0.76	0.53	0.60	0.55	0.67	<b>1.00</b>
chpR - chpL	0.62	0.79	1.12	1.31	0.83	1.15	0.82	1.06	0.99	1.19	0.86	1.06	<b>0.99</b>
chR - chL	0.94	1.23	1.41	1.63	0.99	1.30	1.32	1.49	1.17	1.48	1.08	1.34	<b>0.99</b>
otoR - otoL	1.11	1.39	1.15	1.54	0.89	1.16	0.94	1.37	1.22	1.79	0.94	1.29	<b>1.00</b>
n - gn	1.27	1.85	1.28	1.63	1.05	1.52	1.18	1.59	1.61	2.18	1.19	1.69	<b>1.00</b>
ls - li	0.94	1.14	0.86	1.12	0.89	1.08	0.94	1.14	0.89	1.13	0.68	0.93	<b>1.00</b>
sn - prn	0.36	0.44	0.45	0.55	0.47	0.56	0.40	0.49	0.40	0.50	0.33	0.43	<b>1.00</b>
sn - gn	1.25	1.77	1.19	1.63	1.12	1.46	1.42	1.62	1.35	1.80	1.30	1.52	<b>1.00</b>
<i>Overall</i>	<i>0.75</i>	<i>1.01</i>	<i>0.85</i>	<i>1.12</i>	<i>0.75</i>	<i>1.01</i>	<i>0.84</i>	<i>1.08</i>	<i>0.84</i>	<i>1.10</i>	<i>0.76</i>	<i>0.98</i>	

**Table 15** The reproducibility error of inter-landmark distances in 20 control subject face images according to the angle in degrees clockwise of the camera with respect to the subject's face. The overall mean reproducibility error in each position is 0.85mm or less in all angles tested. The intraclass correlation coefficient (ICC) is above 0.95 for each inter-landmark distance between the different camera angles.

	DSM			
Position	Face3	Eyes3	Nose3	Mouth3
Neutral	8.44	11.75	8.54	5.97
Yaw -5	8.81	11.71	8.35	6.3
Yaw 5	8.76	11.84	8.44	6.15
Roll -5	8.75	11.54	8.38	5.98
Roll 5	8.71	11.72	8.65	6.46
Pitch -5	8.44	11.72	8.51	6.12
Pitch 5	8.48	11.79	8.63	6.35
Significance	$p=0.73$	$p=0.39$	$p=0.77$	$p=0.33$

**Table 16 Median FSD values of twenty control subjects, in seven different camera angles relative to the subject's face. The final row shows that there is no significant difference in FSD values in any camera angle for any of the DSMs as calculated by Friedman's two-way analysis of variance by ranks.**

In the analysis of one subject in greater detail, 125 images were obtained and Figure 32 shows 25 of the images over a range of five angles for both pitch and yaw, with roll kept unchanged from the neutral position. Reproducibility of inter-landmark distances appeared to be worse at angles greater than five degrees in any of the three axes. FSD also increased with angles over five degrees anticlockwise or clockwise. This is shown in one subset in Table 17, in which FSD for the whole face and the overall mean inter-landmark distance error are displayed for the 25 images shown in Figure 32.



*Figure 32 Untextured raw images captured for one control subject with different angles, in degrees clockwise, of pitch and yaw for the camera system. Roll angle was kept at zero from neutral. The images are not to scale but show that different parts of the face are visible at different angles. In particular, the underside of the chin is better captured with a lower pitch angle and a yaw angle close to zero is best for capturing both ears. The tripod mount used to maintain the same face position is also seen in some images but this was outside the face region used for landmarking or dense surface model creation.*

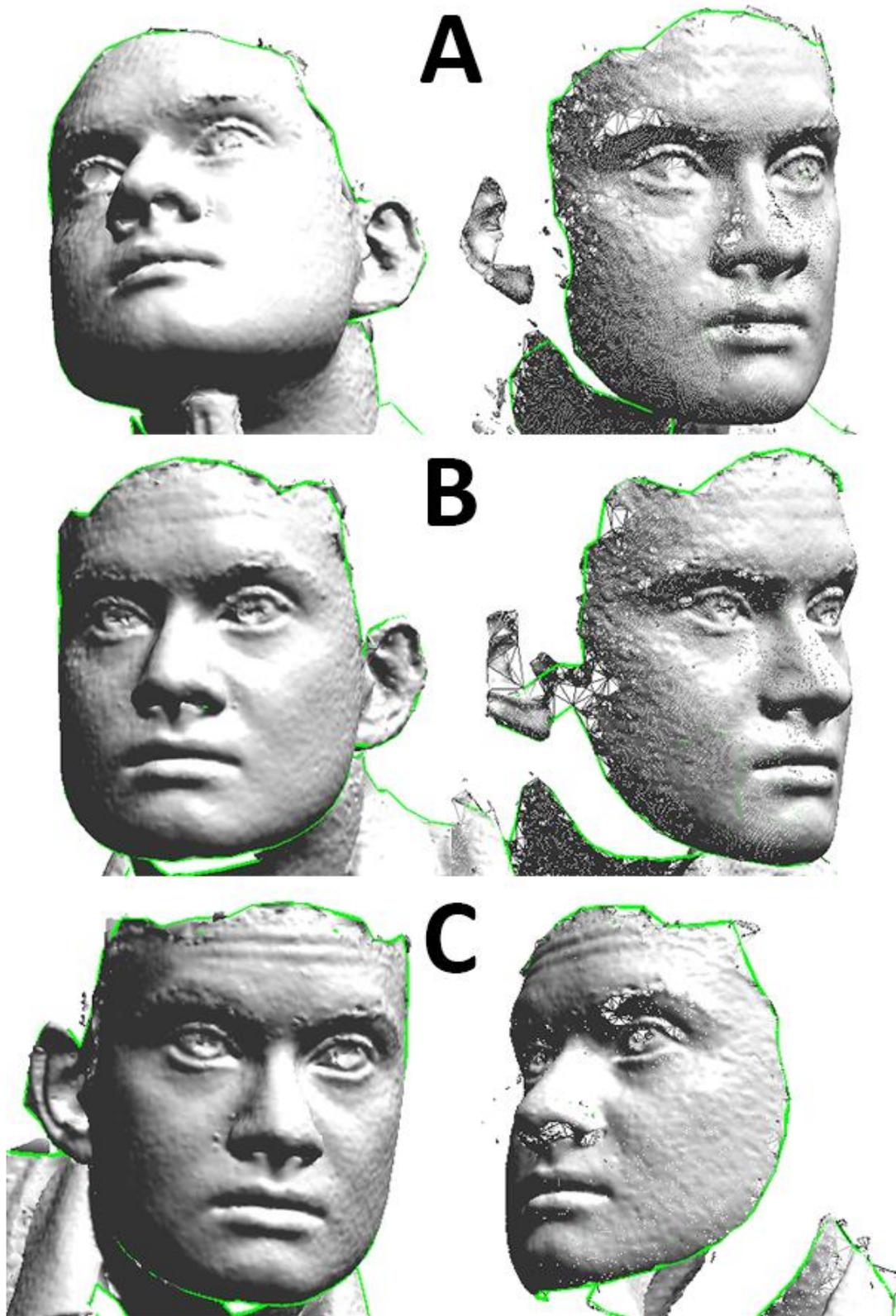
		Yaw (degrees)				
		-10	-5	0	5	10
Pitch (degrees)	-10	11.8	6.5	6.0	6.9	9.6
		15.8	8.7	8.8	9.4	14.5
		7.6	6.3	6.0	7.8	36.8
		7.2	4.4	4.6	5.3	4.9
		0.82	0.32	0.05	0.29	0.33
	-5	8.1	5.9	5.8	6.1	8.7
		9.0	7.6	7.8	8.0	8.2
		6.2	5.7	5.5	6.3	6.8
		4.7	4.9	5.0	4.6	5.1
		0.66	0.70	0.18	0.57	0.62
	0	12.4	6.2	6.5	6.0	8.3
		15.5	8.3	9.2	8.0	9.0
		6.3	5.4	6.2	6.3	5.6
		6.5	4.6	5.1	4.7	4.5
		0.65	0.50	0.00	0.30	0.57
	5	7.8	6.2	6.3	5.6	7.6
		8.8	8.5	9.4	8.4	9.3
		5.4	5.9	6.2	5.5	7.3
		4.9	4.8	5.0	5.2	4.6
		0.45	0.72	0.08	0.32	1.24
10	12.4	8.7	6.6	7.1	10.2	
	8.5	8.0	7.9	7.8	10.7	
	5.6	5.1	5.2	5.2	6.0	
	4.4	4.8	5.0	4.3	4.4	
	0.60	0.06	0.09	0.21	0.47	

**Table 17** Grid of reproducibility errors for 25 face images of one subject taken with different angles of camera pitch and yaw. FSD values for the whole face, periorbital region, perinasal region and perioral region are shown in the top four rows respectively with overall mean inter-landmark distance errors on the lowest row. The values correspond to the images in Figure 32. FSD values over 8 for the whole face are shaded in light grey and values over 10 are shaded in medium grey. FSD appears to increase with greater angles (clockwise or anticlockwise) in yaw. Inter-landmark distance errors also appear to increase with yaw.

The relationship of FSD with each axis was found to be significantly different for differences in pitch (median FSD 6.28 vs 6.34 vs 7.31 at 0, 5, 10 degrees;  $p=0.002$ ;  $n=125$ ) or yaw (median FSD 6.10 vs 6.34 vs 8.47 at 0, 5, 10 degrees;  $p<0.001$ ;  $n=125$ ) but not for roll (median FSD 6.89 vs 6.57 vs 6.69 at 0, 5, 10 degrees;  $p=0.669$ ;  $n=125$ ). The relationship between angle of rotation and FSD does not appear to be linear. There is little difference between 0 and 5 degrees but a larger increase in FSD between 5 and

10 degrees (Figure 36). A similar increase was found with FSDs of the periorbital (Eyes3), perinasal (Nose3) and perioral (Mouth3) DSMs.

In face images with the highest FSD values, visual inspection showed that the camera angles were large enough to prevent complete capture of the face. Figure 33 shows three examples of face images illustrating the loss of detail in various parts of the face that would normally be captured in a neutral position.



**Figure 33** Three examples of face images A, B and C, which had high FSD values. The original image position is shown on the left and the same image is rotated on the right to show missing regions of the face. Image A was taken at pitch=-10, roll=-10, yaw=-10 degrees and misses parts of the right side of the face including the ear. There are also missing areas over the nose surface, the right eyebrow and the right lower lip. Image B was

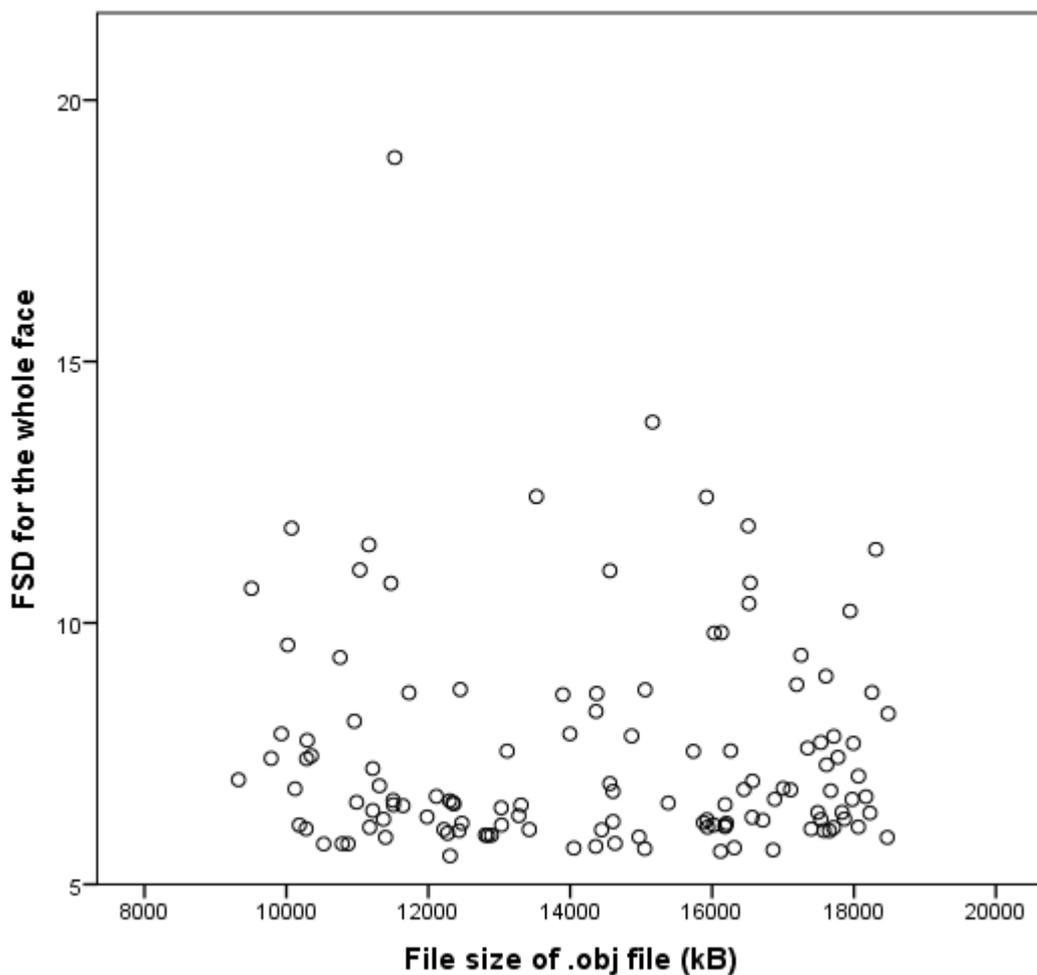
*taken at pitch=10, roll=-5, yaw=-10 degrees and also misses areas of the right side of the face. Image C was taken at pitch=10, roll=5, yaw=10 degrees and shows the left periorbital region and the underside of the nose have missing areas.*

In the three face images, A, B and C, FSD for the whole face was 18.9, 13.8 and 8.82 respectively, significantly higher than the median FSD of 6.52. They were taken at pitch, roll and yaw angles of (-10,-10,-10) degrees for A, (10,-5,-10) degrees for B and (10,5,10) degrees for C. At these angles it appears that there are significantly larger areas of the face that are missing. In face images A and B, the right side of the face is poorly captured. There are areas of the face that are completely missing or in which one can see 'holes' in the face. These 'holes' are large triangular polygons between vertices that represent surface point co-ordinates. If point co-ordinates were not captured in one part of the face, then the resulting triangular polygons are large and visible. For example, in image A, there is missing detail over the right eyebrow, the right lower lip and the right side of the nose. Looking at the original image position, these areas appear to be obscured due to the angle. In image B, the same occurs but to a lesser extent. Image C shows that the underside of the nose and the region under the eyebrows is poorly captured if the camera is not pitched up higher (anticlockwise). This confirms the need to position subjects carefully to capture all areas of the face. In later studies, a check of face coverage was done by immediate visual inspection of each captured 3D face image including rotation of the wireframe image on screen.

#### 3.4.7 IMAGE FILE SIZE AND FSD

Since there are large areas that are missing, it is possible that the images with high FSDs have fewer surface points stored in total and so a smaller image file size. For each face image, all of these surface points are stored in a Wavefront object (.obj) file with tens of thousands of point co-ordinates. The file contains these co-ordinates prefixed by 'v', which are then followed by surface face definitions prefixed by 'f' and comprising references to three or more vertices. All texture and colour information is stored in other files. The overall number of characters and therefore file size of an .obj file is

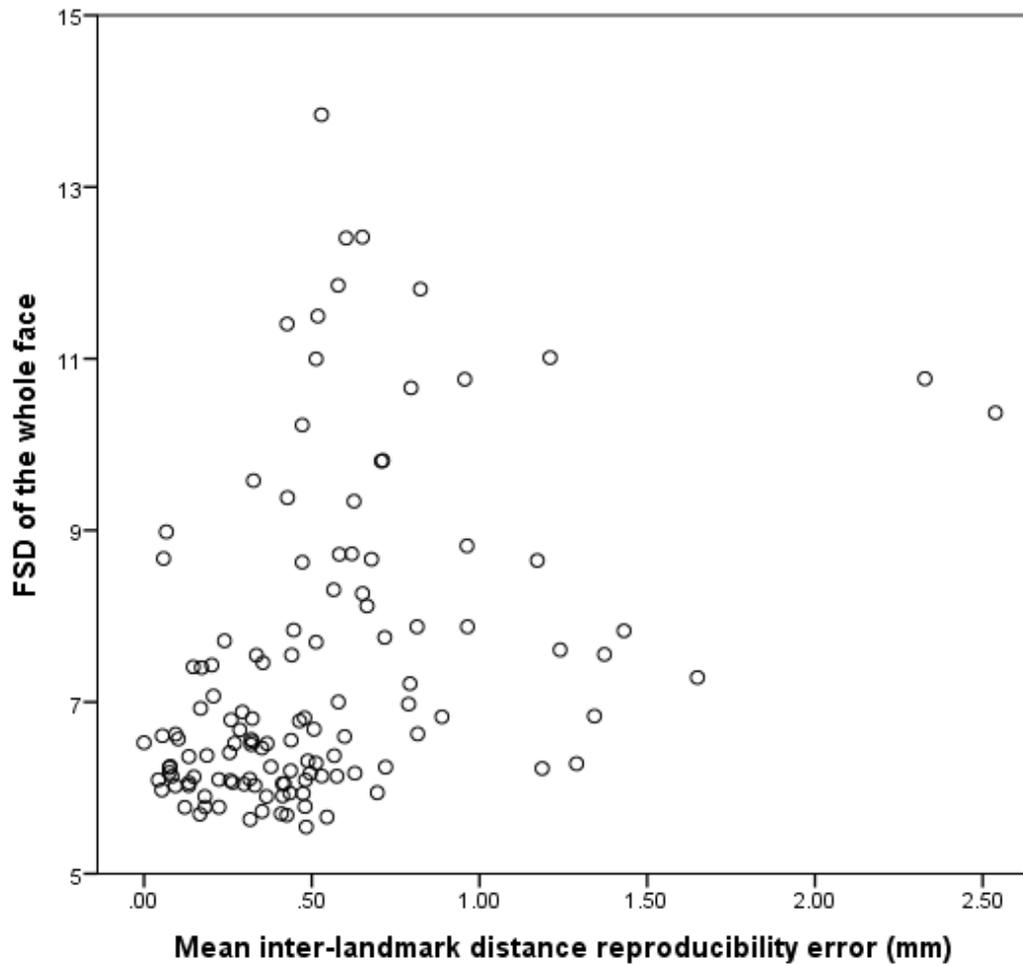
therefore proportional to the number of vertices contained within the file. It was found that there is no relationship of the file size with whole face FSD ( $\rho=0.00$ ;  $p=0.98$ ; Figure 34). This suggests that missing point co-ordinates or lower resolution alone do not explain the greater FSD values seen. A limitation of this exercise is that in images with missing parts of the face, there may be increased inclusion of other extraneous parts in the image. For example, in Figure 33, part of the subject's neck and shirt collar had been included in some images, which may mean the overall number of surface point co-ordinates in the image remained the same as an image in neutral head position.



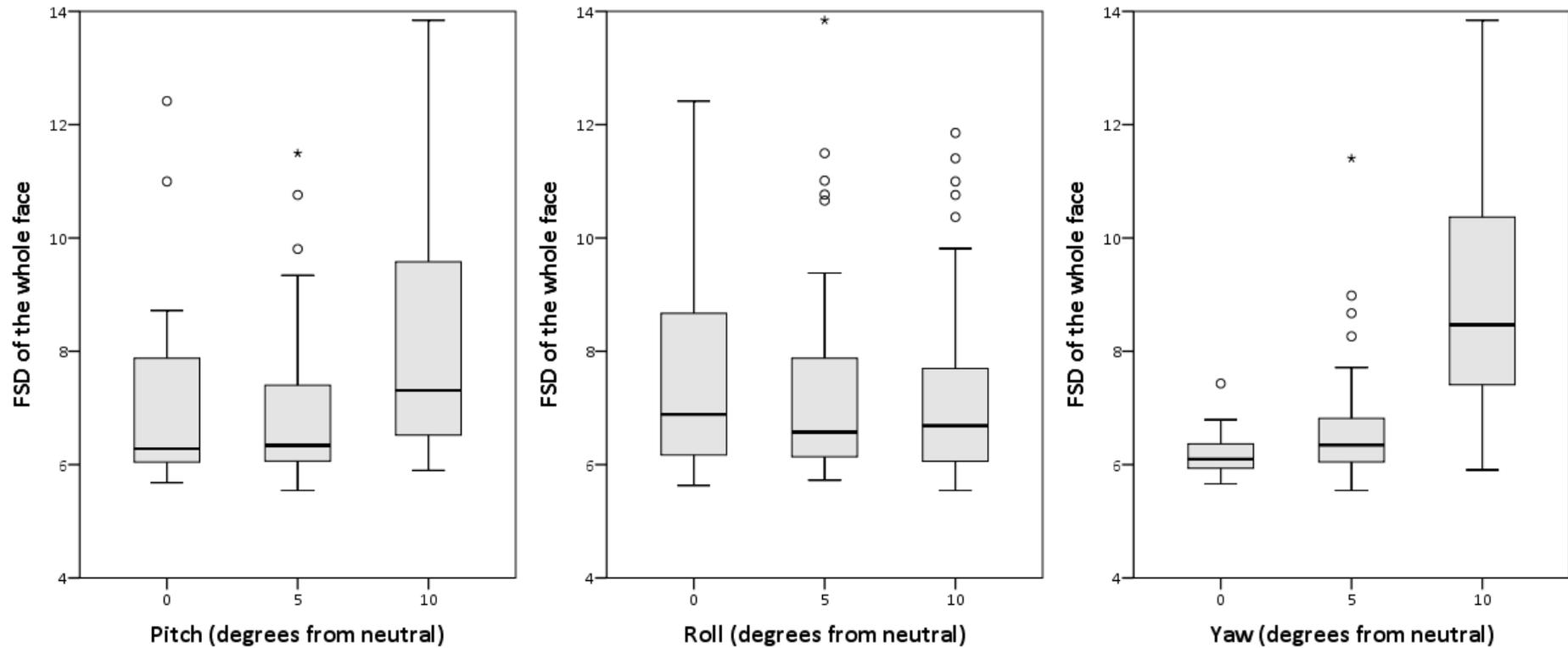
**Figure 34** Scatter plot of file size of a face image's object file size in kilobytes (kB) and that face image's FSD for the whole face, in a set of 125 images of the same subject taken at different angles. There is no correlation between file size and FSD ( $\rho=0.00$ ;  $p=0.98$ ).

### 3.4.8 INTER-LANDMARK DISTANCES AND FSD

For inter-landmark distance error, there was no significant difference due to change in pitch angle (mean error 0.50 vs 0.53 vs 0.50 at 0, 5, 10 degrees;  $p=0.951$ ) or roll angle (mean error 0.43 vs 0.53 vs 0.54 at 0, 5, 10 degrees;  $p=0.551$ ). However, there was a significant difference in error with change in the angle of yaw (mean error 0.21 vs 0.46 vs 0.71 at 0, 5, 10 degrees;  $p<0.001$ ). Given that landmarking errors appeared to increase with change in camera yaw angle, I considered that the landmarking errors alone could explain the changes seen in FSD at different camera angles. Correlation between FSD and inter-landmark distance errors in all 125 images was weak though positive and significant ( $\rho=0.46$ ;  $p=6.45 \times 10^{-8}$ ; Figure 35). Therefore, landmarking error alone does not seem to explain the reduced reproducibility of FSD with different camera angles. Other possible errors are those to do with repeat image capture: calibration, facial expression and facial position.



*Figure 35 Relationship between FSD of the whole face and the mean inter-landmark distance reproducibility error for 125 faces of the same control subject. Correlation was weakly positive (Spearman's rank correlation coefficient  $\rho=0.46$ ;  $p=6.45 \times e-8$ )*



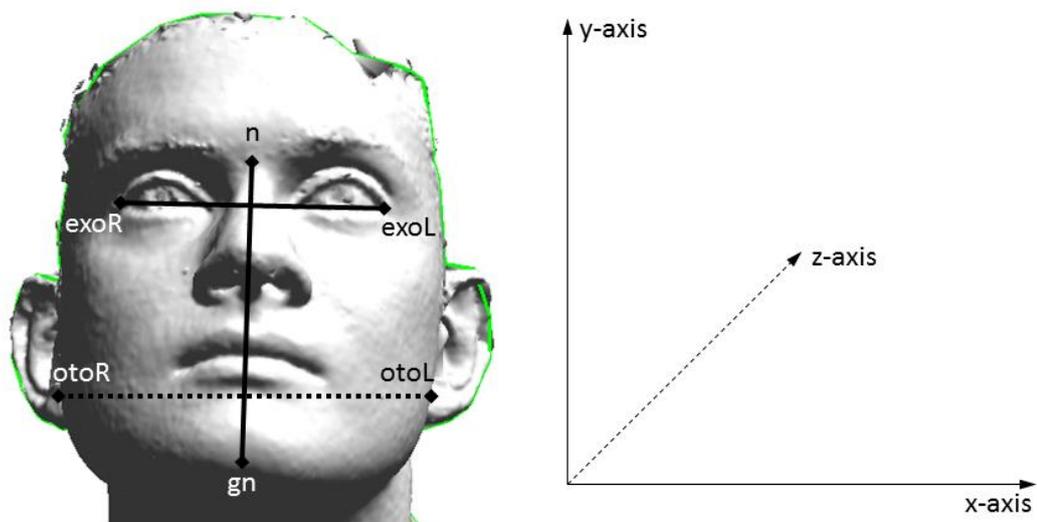
**Figure 36** Graphs of the distribution of FSD for the whole face for different camera angles from the neutral position. There is no significant difference in FSD for camera roll angles up to 10 degrees (median FSD 6.89 vs 6.57 vs 6.69 at 0, 5, 10 degrees;  $p=0.669$ ;  $n=125$ ) but there is a significant difference in FSD for changes in camera pitch angle (median FSD 6.28 vs 6.34 vs 7.31 at 0, 5, 10 degrees;  $p=0.002$ ;  $n=125$ ) and camera yaw angle (median FSD 6.10 vs 6.34 vs 8.47 at 0, 5, 10 degrees;  $p<0.001$ ;  $n=125$ ). For angles, only the magnitude is considered so clockwise and anticlockwise angles are treated the same.

### 3.4.9 DETERMINATION OF HEAD POSITION USING INTER-LANDMARK DISTANCES

Given that FSD may be affected by camera angles, I sought to see if the angle of the face image could be reliably determined in any of the three axes from the raw landmark co-ordinates. Absolute landmark co-ordinates will vary as people will sit in different postures, different heights and different lengths relative to the camera for each image. On the other hand, inter-landmark comparisons in specific axes may be able to provide a reference.

Within each object file, the manufacturer states that the Canfield Vectra CR10 system stores points in a 3D frame where the x-axis is from left to right, the y-axis is from down to up and the z-axis is the distance away from the camera (Figure 37). I chose three inter-landmark distances to check the 3D frame and see how they varied with camera rotation. Biocular width, the distance between the exocanthions, has been used before as the most reliable horizontal line on face surfaces, with little displacement in the y-axis and z-axis (Maal et al., 2010; Plooij et al., 2009; Ras et al., 1996). Facial width, the distance between the otobasions, was used as a second potential horizontal line. The facial height, as judged from nasion to gnathion, runs through the midline of the face and was a third distance chosen for the least displacement in the x-axis. Midline face landmarks have also been used previously to determine the vertical plane (Toma et al., 2009). Note that the above studies have also used a straight line from exocanthion to the most superior point of the ear to define the horizontal plane. This assumes that ears are set at a normal height and in dysmorphology, it is well-known that low-set ears are a feature of many disorders. I found no dysmorphological conditions that cause a difference in the vertical position of one eye or ear compared to the other, although there may be natural variation.

Next the displacement of each of the three lines in each of the x, y and z axes was calculated for the different camera angles. For camera pitch, the greatest change was in the z-axis of the nasion to gnathion line. For camera roll and yaw, the greatest changes were in the y-axis and z-axis respectively, of both of the chosen horizontal lines. The biocular width line was chosen instead of facial width as this has been used before in studies due to its reproducibility. The mean displacement and its standard deviation are shown in Table 18. For the study of one subject at five angles per axis of rotation, there was a strong relationship between axis displacement and camera angle. Simple linear regression was used to show that 92% of the variation in the z-axis of face height was explained by camera pitch ( $r^2=0.92$ ;  $p= 1.12 \times 10^{-69}$ ,  $n=125$ ). Similarly, 94% and 95% of the variation in y-axis and z-axis of biocular width was explained by camera roll ( $r^2=0.94$ ;  $p= 5.61 \times 10^{-78}$ ,  $n=125$ ) and yaw ( $r^2=0.95$ ;  $p= 3.21 \times 10^{-80}$ ,  $n=125$ ).



**Figure 37** Viewing from the Canfield Vectra CR10 camera, the z-axis is the optic axis of the camera, equivalent to the direction perpendicular to the plane of the four camera lenses. The x-axis and y-axis are horizontal and vertical, relative to the camera, in a plane parallel to the plane of the camera lenses. I chose three of the longest inter-landmark distances to assess if they could help determine facial angle relative to the camera reliably. These distances were right exocanthion to left exocanthion (exoR – exoL, biocular width), nasion to gnathion (n – gn, facial height) and right otobasion inferius to left otobasion inferius (otoR – otoL, facial width).

Rotation	Angle	Distance	Axis	Within one subject		Within 20 subjects	
				Mean (mm)	SD (mm)	Mean (mm)	SD (mm)
Pitch	-10	n-gn	z	-60.3	3.1	-	-
Pitch	-5	n-gn	z	-48.0	3.8	9.2	6.8
Pitch	0	n-gn	z	-38.2	4.4	15.0	10.1
Pitch	5	n-gn	z	-31.3	5.2	17.7	9.2
Pitch	10	n-gn	z	-14.3	5.5	-	-
Roll	-10	exoR-exoL	y	-16.6	2.0	-	-
Roll	-5	exoR-exoL	y	-9.0	2.9	-1.5	4.5
Roll	0	exoR-exoL	y	-1.7	3.2	-1.5	4.8
Roll	5	exoR-exoL	y	6.1	3.5	-0.4	5.8
Roll	10	exoR-exoL	y	12.7	3.9	-	-
Yaw	-10	exoR-exoL	z	-16.6	1.7	-	-
Yaw	-5	exoR-exoL	z	-10.1	1.9	-1.7	3.7
Yaw	0	exoR-exoL	z	-2.4	2.4	1.6	3.2
Yaw	5	exoR-exoL	z	3.9	2.4	5.3	4.4
Yaw	10	exoR-exoL	z	12.2	2.6	-	-

**Table 18 Displacement in one selected axis of an inter-landmark distance showing the relationship with camera rotation. For the study of one subject and the study of 20 subjects, mean displacement in the selected axis increased with an increase in the camera angle. There was greater standard deviation between subjects than in the same subject. This suggests that though displacement between landmarks can be used to estimate the camera angle relative to the face, it is only helpful for repeated images of the same face.**

Using the same measures of displacement in the set of 20 control subjects, the change in axis for all three distances is much less predictable. Simple linear regression shows that only part of the displacement could be predicted by camera angle. This was true for camera pitch ( $r^2=0.10$ ;  $p= 0.004$ ,  $n=80$ ), camera roll ( $r^2=0.01$ ;  $p= 0.46$ ,  $n=80$ ) and camera yaw ( $r^2=0.32$ ;  $p= 5.22 \times 10^{-8}$ ,  $n=80$ ). Therefore, change in displacement between suitable face landmarks is not a useful measurement to quantify the angle of the camera relative to the face. Controls were not assessed as fully as for one individual subject so it is not known if including camera angles 10 degrees from neutral would have increased the accuracy.

In summary, although detailed analysis in repeated face images of one subject suggested that head position and angle could be inferred, using multiple subjects it was

not possible to reliably assess head position from a 3D face image. Thus, if head position is to be controlled for, then it has to be done at the time of image capture and cannot be checked retrospectively from captured images. However, there is no method of accurately fixing head position during image capture without interfering with image acquisition. I used an adjustable swivel chair and rotated the chair so that the head approximately faced the vertical midline of the camera (approximating yaw=0°) on visual inspection. I also used a visual target above the camera and adjusted the height of the chair so that eye level approximated the top of the camera, again on visual inspection from the side of the subject (approximating pitch=0°). Roll could not be reliably assessed prior to image capture, but has been noted to make no significant difference to FSD for variations from -10° to 10°. Further studies are needed to look at ways of fixing head position for accurate stereophotogrammetry.

#### 3.4.10 FACIAL EXPRESSIONS

Facial expressions are likely to affect face shape, including landmark positions, and thus FSD. Three facial expressions were assessed in the twenty control subjects. Inter-landmark distances as measured on these face images were compared to the distances on the face image with a neutral expression and the mean error for each distance in all subjects is shown in Table 19. The same twelve inter-landmark distances used for inter-operator reproducibility are used here and can be seen in Figure 29, Figure 30 and Figure 31.

The repeat images taken in a neutral expression after re-calibration are shown on the right-hand column as a comparison. Data were normally distributed and a one-way ANOVA was used to assess if there was any significant difference in any distance with four different facial expressions. Distances between the exocanthions, endocanthions and otobasions were not affected significantly by any facial expression. All other distances were significantly altered.

	Pucker		Smile		Mouth open		Calibration		Significance
	Mean (mm)	SD (mm)	Mean (mm)	SD (mm)	Mean (mm)	SD (mm)	Mean (mm)	SD (mm)	One-way ANOVA
exoR - exoL	0.15	0.81	0.19	1.71	0.06	1.31	0.11	0.47	$p=1.00$
endoR - endoL	0.14	0.92	0.65	1.95	0.09	1.12	0.23	0.70	$p=0.88$
psupR - pinfR	0.94	1.03	2.84	1.79	0.62	1.53	0.25	0.63	$p=5.5 \times e^{-8}$
psupL - pinfL	1.01	1.17	3.07	1.84	0.86	1.65	0.31	0.76	$p=1.5 \times e^{-7}$
alaR - alaL	1.42	1.64	4.00	1.72	1.17	1.77	0.21	0.59	$p=0.0003$
chpR - chpL	3.42	2.23	3.05	2.53	0.79	2.29	0.15	0.82	$p=0.000001$
chR - chL	12.2	7.14	14.3	4.95	0.80	5.98	0.11	1.73	$p=4.1 \times e^{-19}$
ls - li	0.40	1.62	2.62	2.23	0.42	4.10	0.05	1.32	$p=6.6 \times e^{-36}$
otoR - otoL	2.12	5.72	2.20	3.47	38.1	6.34	0.31	1.35	$p=0.88$
n - gn	8.60	5.61	7.45	5.42	45.2	7.47	0.17	1.70	$p=1.9 \times e^{-22}$
sn - prn	1.97	2.17	0.66	1.18	0.22	0.86	0.09	0.42	$p=0.002$
sn - gn	2.34	5.80	3.72	3.25	41.1	6.27	0.16	1.82	$p=1.1 \times e^{-29}$

**Table 19** The inter-landmark distance reproducibility error in 20 control subject face images according to facial expressions. Expressions included lip puckering (Pucker), smiling with lips open (Smile) and opening the mouth wide (Mouth open). The errors are compared to the error found during re-calibration in a neutral expression. Analysis of variance (ANOVA) testing shows that there was a significant difference in error for all distances except those using exocanthion, endocanthion and otobasion.

Tukey's method was used as a post-hoc test to explore which face expressions affected each distance. Distances are shown in Table 20 along with face expression(s) that significantly alter their length. This shows that smiling, puckering lips or opening the mouth all increase the distance between the lips in the midline. In addition, the different facial expressions affect all of the other expressions in different ways. Smiling increases the width of the lips, philtrum and the nostrils but reduces the height of the palpebral fissure. Lip puckering reduces the width of the lips and philtrum. Lip puckering is the only expression that also reduces nasal protrusion. Opening the mouth affects all distances measured across the lips including lip height, face height and lower face height. Mouth opening also increases philtrum width but not lip width.

Distance	Expressions reducing distance	Expressions increasing distance
exoR - exoL	Nil	Nil
endoR - endoL	Nil	Nil
psupR - pinfR	Smile	Nil
psupL - pinfL	Smile	Nil
alaR - alaL	Nil	Smile
chpR - chpL	Pucker	Mouth open, Smile
chR - chL	Pucker	Smile
ls - li	Nil	Mouth open, Pucker, Smile
otoR - otoL	Nil	Nil
n - gn	Nil	Mouth open
sn - prn	Pucker	Nil
sn - gn	Nil	Mouth open

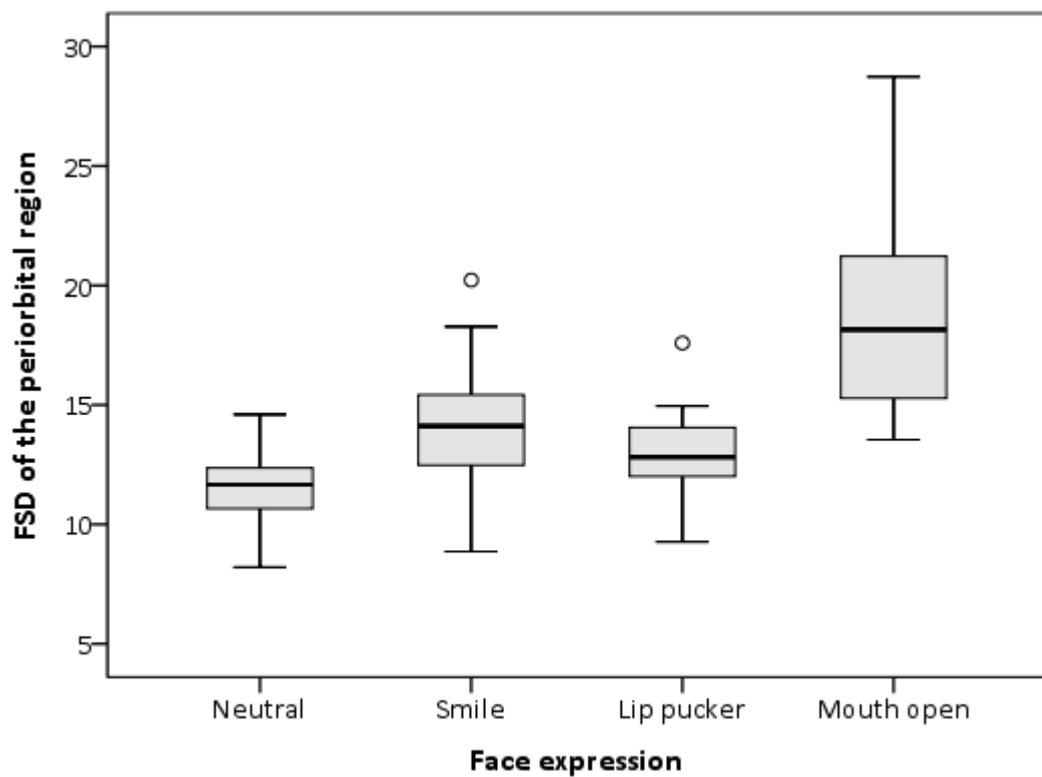
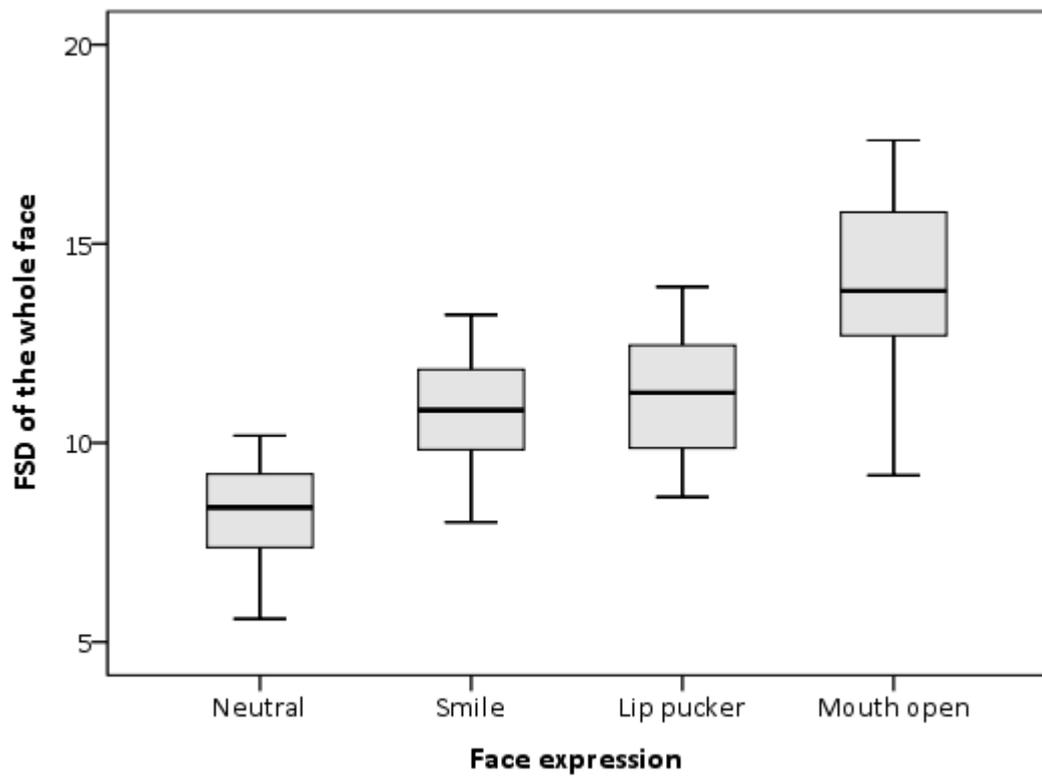
**Table 20 Summary of which facial expressions significantly reduce or lengthen each inter-landmark distance using the post-hoc Tukey's test.**

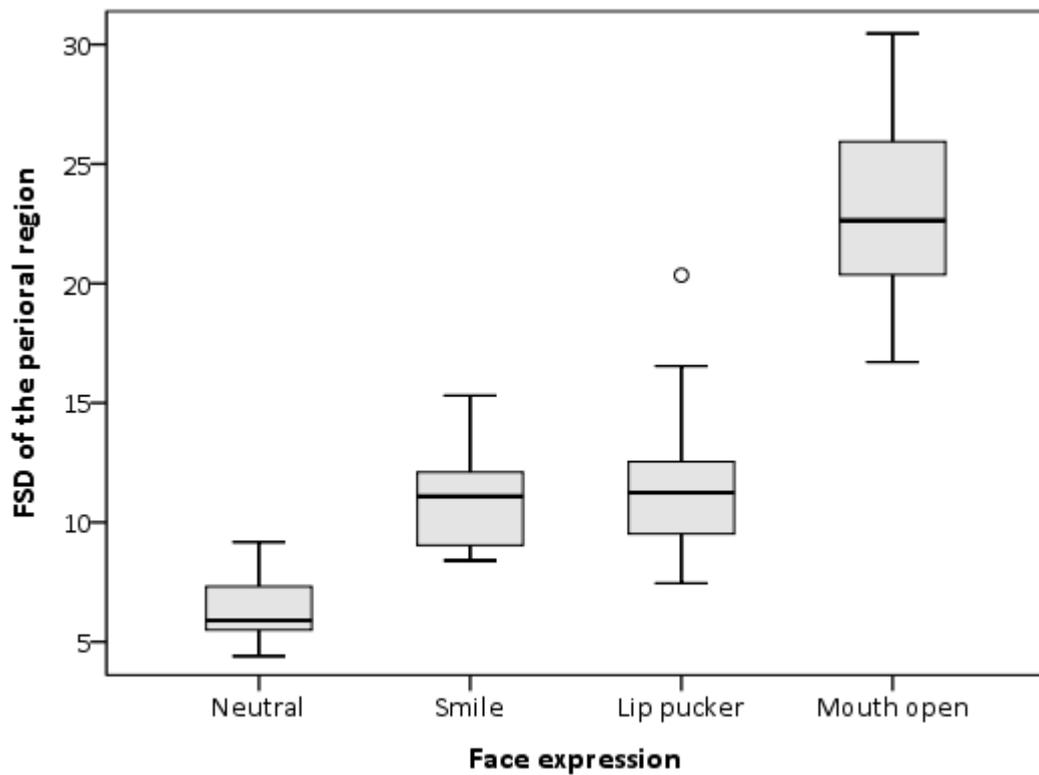
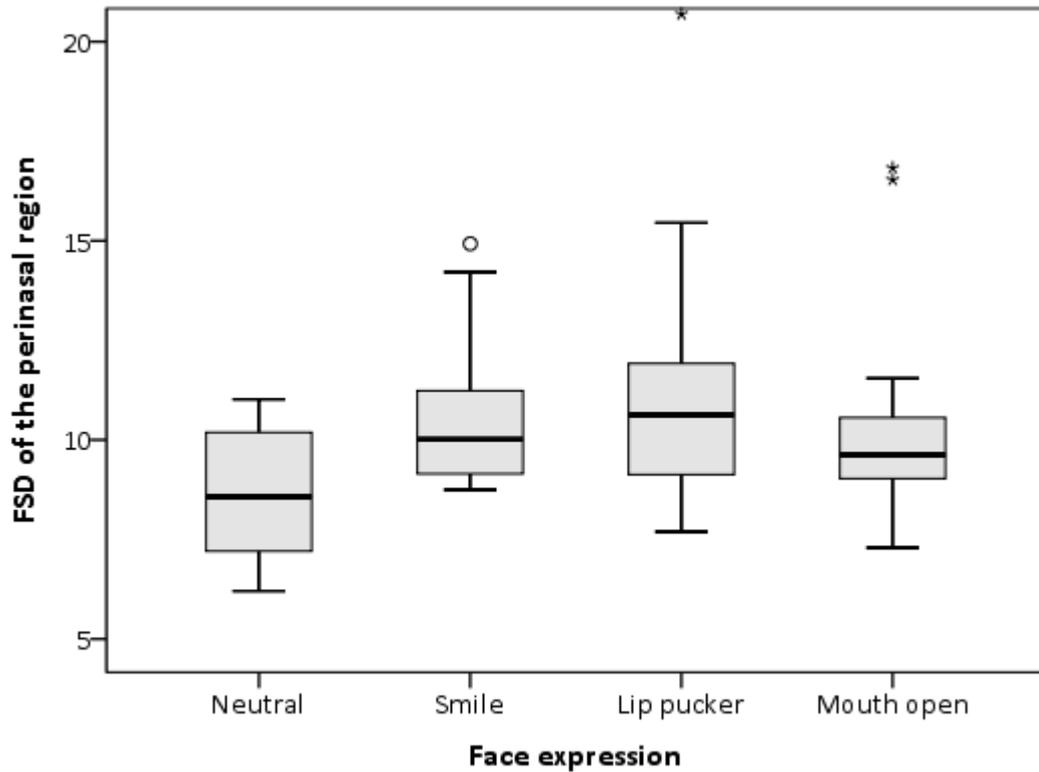
FSD values for the twenty control subjects in all four DSMs were significantly different between the four facial expressions (Table 21).

DSM	Facial expression				Significance
	Neutral	Smile	Pucker	Mouth open	
Face3	8.38	10.8	11.3	13.8	$p<0.001$
Eyes3	11.7	14.1	12.8	18.2	$p<0.001$
Nose3	8.57	10	10.6	9.62	$p<0.001$
Mouth3	5.9	11.1	11.2	22.6	$p<0.001$

**Table 21 The effect of different facial expression on the mean FSD in the twenty subjects. Significance was calculated using a related samples Friedman's two-way analysis of variance.**

FSDs were highest with the mouth open for the whole face, periorbital and perioral DSMs. For the perinasal DSM, lip puckering and smiling caused the biggest increase in FSD. The distribution of FSD is shown in more detail for each of the four DSMs in Figure 38. It can be observed that the periorbital DSM showed the highest FSDs with the mouth open even though no distance involving periorbital landmarks was significantly altered by mouth opening. Therefore, facial expressions that are not neutral may significantly alter FSD and need to be controlled for. In later studies, this was done by repeat instructions to maintain a neutral expression, as noted in Chapter 2, and visual inspection of the face itself and the 3D face image immediately after capture.





**Figure 38** Box plots of the distribution of FSD with four facial expressions: Neutral, smiling with lips open, lips puckered and mouth wide open. FSD is shown for DSMs of the whole face (Face3), periorbital region (Eyes3), perinasal region (Nose3) and perioral region (Mouth3). FSD varied significantly with facial expression in all four DSMs.

### 3.5 DISCUSSION

In this chapter, I have shown that reproducibility of stereophotogrammetry using the Canfield Vectra CR10 camera is comparable to previous studies. For inter-landmark distances (which are typically in the range 10-120mm), overall inter-operator reproducibility error was 0.71mm in controls. Overall intra-operator reproducibility error was 0.34mm in people with epilepsy. The reproducibility error is comparable to other studies of living subjects, which show mean errors of 0.26mm to 1.3mm for intra-operator error and 0.29mm to 2.8mm for inter-operator error (Table 5). Intraclass correlation coefficients were between 0.94 and 1.00 for these measurements. The intraclass correlation coefficients are better than those previously obtained (Dindaroğlu et al., 2015; Othman et al., 2013).

For point co-ordinates of landmarks themselves, the mean reproducibility error was 0.94mm (intra-operator) and 1.31mm (inter-operator). This error is comparable to that seen in one study of point co-ordinate reproducibility in living subjects (Gwilliam et al., 2006), but greater than in another (Plooij et al., 2009). One other study also showed greater reproducibility but the authors used mannequin faces (Ayoub et al., 2003). This also illustrates one of the drawbacks of only using inter-landmark distances for reproducibility error. Absolute errors in the position of two landmarks may be cancelled out in inter-landmark distances (e.g. if gnathion was 5mm superior and nasion was 5mm superior, then the inter-landmark distance from nasion to gnathion would still appear to be the same).

To my knowledge, this is the first study that investigated the effect of camera calibration or head positioning on landmark placement reproducibility on living subjects. Two studies have looked at head position for mannequin faces previously and

only one of these looked at head positions in three dimensions (Lee et al., 2004; Lübbers et al., 2010). For images taken before and after calibration, the mean intra-operator reproducibility error between inter-landmark distances was 0.18mm. This error is a combination of calibration itself, image acquisition error and landmarking error, since each image was acquired and landmarked separately (by Vamsi Chinthapalli). In fact, reproducibility error was less than the reproducibility error for patients' images that were landmarked twice on the *same* image (0.18mm vs 0.34mm). This suggests that landmarking error is greater for patients with epilepsy or that landmarking precision improved with experience. The operator had landmarked an additional 1000 images between the first landmarking reproducibility study and this second study of calibration and landmarking reproducibility.

Head positioning changes worsened landmark placement reproducibility. When analysing inter-landmark distances, there was a greater overall intra-operator reproducibility error of 0.75-0.84mm, compared to 0.34mm in the neutral position. The reason for this may be due to poorer image quality in regions with landmarks of interest. With camera yaw, I have shown that there are missing areas of the face on images and this would affect landmarks closest to the periphery of the face. These landmarks include otobasions and gnathion and distances using these landmarks were the ones showing the greatest error.

Facial expression, as expected, significantly affected landmark placement. Similar to previous studies, opening the mouth causes the greatest change overall in inter-landmark distances. There is variation in the extent of change seen in these distances. Distances between the eyes and the ears were not affected by facial expressions. The distances using lip landmarks were most affected and those involving the nose were also affected to a lesser degree.

The reproducibility of individual landmarks is comparable to previous findings, in which the majority of landmarks show a mean discrepancy of under 1mm. The least reproducible landmarks (>1mm error) comprised those with no clear surface features or boundaries to guide placement in all subjects, and comprised cheilion, nasion, gnathion, alari nasi, and otobasion inferius. Cheilion is at the corner of the mouth, but with the lips closed, a crease is often present here and it is difficult to judge exactly where the vermilion lip border is. Nasion, gnathion and alare nasi are all placed on the face surface where there are no discrete changes in shape contour or skin colour. Otobasion inferius, at the junction of the earlobe and the cheek, can be straightforward to place in some subjects but there are subjects who have attached earlobes, again with no obvious change in contour. Also, accurate image detail for otobasion is lacking when subjects are laterally rotated, which may also reduce reproducibility. All of these landmarks have been found to be poorly reproducible in multiple earlier studies too, except for otobasion inferius (Table 6), which was only used in one previous study (Gwilliam et al., 2006).

For FSD, the overall reproducibility error appears to be low. For example, using the whole face DSM (Face3), FSD for all 163 subjects varied from 6.2 to 31.3 with a median absolute difference of 1.0 unit (Chapter 4). Compared to this, the median change in FSD in 20 control subjects who had undergone repeat camera calibration, repeat image acquisition and repeat landmarking, was 0.1 units. In different DSMs created specifically to check for landmark reproducibility, FSDs would be different because the principal components created in each DSM are unique. However, there was a strong positive correlation between the original DSM FSD and the repeat DSM FSD for the whole face, the periorbital region or the perinasal region ( $\rho=0.90-0.94$ ;  $\rho^2=0.81-0.88$ ).

FSD appears to be more sensitive to changes in head positioning and facial expressions though. Camera pitch and yaw (but not roll) between 5 and 10 degrees in either

direction significantly affected FSD in all DSMs. Yaw is equivalent to a subject laterally rotating his or her head left or right and pitch is equivalent to a subject flexing and extending their head forwards and backwards. Changing head position affects the coverage of the face surface with missing areas seen at extreme angles (10 degrees in any direction). However, the image file size analysis suggests the change in FSD is not simply due to missing point co-ordinate data alone. A limitation of the head positioning analysis is that the camera was changing position not the subject, when in practice the camera would actually be fixed in position. Unfortunately, it was not practical to have subjects turning their head to specified degree intervals in three axes and this may be why no other studies have attempted to assess head position in living subjects. Using camera rotation means that I may be missing changes to landmark position that occur when subjects have a different head position. One hypothetical example is if a subject extends her neck so that her head is angled backwards, then her eyes will be looking 'downwards' towards the camera and this may give the appearance of ptosis. Another part of the head positioning study involved attempting to deduce the angle of head position from a 3D face surface image in the set of 20 control subjects. With repeated images in different positions for one subject, it was possible to estimate the angle using various inter-landmark distances, but unfortunately it could not be extrapolated to different subjects. Therefore, currently the only way of assessing head position was by visual assessment of the subject at the time of image capture. This is not ideal and other methods to assess head position are needed, given its potential impact on reproducibility error. This impact is not just specific to FSD but also to landmark placement. On the other hand, in repeated images of control subjects in the neutral position, as judged by visual inspection, there was good reproducibility. This suggests visual assessment of head position may be adequate at present.

As with head positioning, FSD is affected significantly by facial expressions in all DSMs. Even DSMs of parts of the face have to be used with caution in the presence of certain

expressions that may not be expected to affect those parts. For example, opening the mouth appears to increase FSD of the periorbital region despite little change in landmarks in that region. This finding also suggests that changes in face expression are not always captured by landmark-based anthropometry despite dense surface modelling detecting a difference. For example, lip puckering or opening the mouth both increased FSD for the periorbital region (Eyes3) DSM. However, lip puckering and opening the mouth did not significantly alter inter-landmark distances using endocanthion, exocanthion, palpebrale superius and palpebrale inferius. The increased sensitivity of FSD compared to landmarks may be expected given that FSD takes into account the whole shape of the periorbital region, and not just a number of distance measurements between eight landmarks around the eye. In later chapters, it is shown that there are some subjects in whom the lips were apart. Based on this reproducibility data, such patients were excluded in sensitivity analyses to avoid confounding. No subject smiled or puckered their lips in later chapters. In these chapters, if a high FSD was obtained then the image was visually inspected first for altered head position or face expression. This principle may be necessary for future studies too.

The use of landmarks is debatable too, as landmarking is the major operator-dependent step in the creation of DSMs and calculation of FSD. As mentioned earlier, landmark reproducibility has long been known to be dependent on definitions used and anatomy training, even in older studies using direct anthropometry. The use of DSMs means that the whole face surface is analysed without reliance only on landmarks. However, landmarks are needed for the thin plate spline warping and dense correspondence creation steps. Manually annotated landmarks allow computers to interpolate all other corresponding points. Any landmarks can be used for this but they need to try to encompass the face. If, say, all landmarks were those on the nose and lips alone, then a small error in landmark placement here would affect thin plate spline and dense correspondence more in the periphery of the face.

Advances in computation mean that new techniques are possible to improve the reproducibility of landmarks. There are now algorithms that can identify landmarks on a face surface using a maximum curvature function or by determining intersections of curves (Katina et al., 2016). Already, one technique for automated placement of landmarks is robust enough to work on incomplete face surface images (Sukno et al., 2015). Incomplete face surface images occurred in the above studies, particularly in certain orientations, and previously the inability to automate landmark placement in incomplete images had been a major hindrance. With automation, it may be possible to increase precision and accuracy in FSD further.

One limitation of this study is that stereophotogrammetry was not compared to another method of landmark placement. This means one cannot determine the accuracy of absolute landmark positions from the current analysis. However, this should not affect the assessment of reproducibility. Reproducibility is of more importance as landmarks need to correspond as closely as possible for dense surface modelling and FSD calculation. It means that absolute comparisons of landmark positions or inter-landmark distances should be performed cautiously until accuracy is known. Another limitation is that camera position was changed during the analysis, not subject position or orientation. It could be argued that the effect of gravity on the face has been missed because of this. However, I do not believe this is a significant source of error. It was more important to accurately measure the angle between the camera system and the face to quantify errors and there was no easy method to adjust and maintain a subject's head at a set angle without compromising the 3D image or introducing further errors.

### 3.6 CONCLUSION

Landmark placement using the Canfield Vectra CR10 stereophotogrammetry camera system and in-house software gives comparable intra-operator and inter-operator

reproducibility to previous studies. FSD is reproducible in control human subjects and those with epilepsy in repeated images in the neutral position, with or without camera calibration. Head positioning and facial expressions are found to decrease reproducibility of landmark placement and FSD. These factors need to be considered during image capture. As noted above, this was performed largely using careful instructions and visual cues, adjusting the position of the face indirectly using the subject's chair, and visual inspection of the face and 3D face image after capture.

## CHAPTER 4: PATHOGENIC STRUCTURAL VARIANTS AND FACE SHAPE

### 4.1 INTRODUCTION

The association between chromosomal disorders, facial dysmorphism and epilepsy is well described (Goodman, 1977; Jones, 2006; Singh et al., 2002). Such chromosomal disorders comprise deletions, duplications, inversions and translocations of parts of a chromosome. These may be termed structural variants (SVs) and of these, deletions and duplications are now more commonly grouped together as copy number variants (CNVs) as discussed in Chapter 1. Given that brain development and face development are known to influence each other, it is perhaps not surprising that there is such an association between disorders of face shape and epilepsy, a disorder of the brain.

Large structural variants, which are those that could be detected using karyotyping (approximately >5 megabases of DNA), were previously thought to be only responsible for rare epilepsy syndromes associated with intellectual disability and dysmorphism, such as Wolf-Hirschhorn syndrome (Singh et al., 2002).

Through the use of array comparative genomic hybridization (aCGH) and single nucleotide polymorphism (SNP) microarrays, smaller structural variants (<5 megabases of DNA) are now recognised as an important cause of neuropsychiatric disorders, including epilepsy. These technologies are able to detect structural variants over 25 kilobases in size reliably (Alkan et al., 2011). Indeed, SVs are collectively the commonest genetic abnormality in epilepsy, occurring in up to 10% of people with epilepsy in selected populations. Both recurrent SVs and rare SVs in epilepsy appear to be associated with the presence of facial dysmorphism (Helbig et al., 2014; van Bon et al., 2009).

Stereophotogrammetry and dense surface modelling are reproducible and able to discriminate differences between different types of facial dysmorphism as well as

identifying facial features not recognised on visual inspection alone (Hammond et al., 2012a, 2005). They have already been used in syndromes associated with epilepsy including Smith-Magenis syndrome and Wolf-Hirschhorn syndrome.

Facial shape analysis may therefore be able to help identify patients with underlying recurrent or rare pathogenic SVs, particularly in those with epilepsy.

## 4.2 HYPOTHESES

In a group of people with epilepsy, I hypothesise that stereophotogrammetry and dense surface models can distinguish between those who do and those who do not have pathogenic SVs. I also hypothesise that this remains true after excluding the effect of several potential confounders: individual outliers, age differences, facial injury, face expression, anti-epileptic drug use and presence of intellectual disability.

## 4.3 METHODS

### 4.3.1 RECRUITMENT, IMAGE CAPTURE AND FACE SHAPE ANALYSIS

Patients with epilepsy were recruited as described in Chapter 2. Two groups were analysed in turn – a training set (Group 1) and then a validation set (Group 2). Patients with known structural variants were invited to participate from three European institutions for this study to increase the number of people with epilepsy and SVs. Image capture, landmarking, dense surface model creation and calculation of Face Shape Difference (FSD) was performed as described in Chapter 2. Additional data collection is described below.

### 4.3.2 CLASSIFICATION OF BRAIN IMAGING, INTELLECTUAL DISABILITY AND DRUG HISTORY

Brain MRI scan findings were collected from imaging reports in the clinical record or on a clinical imaging database. Imaging reports were produced by experienced neuroradiologists. They were classified into three categories: normal, normal with

incidental findings and abnormal. 'Normal with incidental findings' was used to describe abnormalities that do not contribute to epilepsy, but nonetheless may be associated with underlying SVs. They comprised mild cortical atrophy, non-specific white matter lesions, and in one case a frontal lobe arachnoid cyst.

Intellectual disability was determined from neuropsychology reports, full-scale or verbal intelligence quotient (IQ) scores, and clinical documentation of level of functioning in daily activities (Salvador-Carulla et al., 2011). Because of the difficulty of accurate retrospective assessment of intellectual disability and the particular difficulty in identifying borderline intellectual functioning (Wieland et al., 2014), especially in cases without formal neuropsychometry, only three categories were used – Normal/Mild, Moderate, or Severe/Profound – applying the same definitions as in the International Classification of Diseases, 10th Revision, by the World Health Organization. For adults, the earliest available neuropsychometric and clinical records were used to minimise confounding from potential effects on intellectual performance of chronic epilepsy, medical and surgical treatments, and neurodegeneration.

Anti-epileptic drug use history was obtained from clinical records. All available records were inspected to check for the use of anti-epileptic drugs, recorded as 'Yes' or 'No'. This approach was limited by the availability of clinical records.

#### 4.3.3 CLASSIFICATION OF DYSMORPHISM, FACIAL INJURIES AND EXPRESSION

For all patients in group 1, four physicians specialising in adult neurology (Isabel Parees, Jasper Morrow, Jan Novy, Tabish Saiffee) were asked to inspect each 3D face image. They were blinded to all clinical details except age and sex. They were asked to decide if there was facial dysmorphism, and were allowed to freely rotate the 3D image to view it from different angles. All images were classified by them as 'Dysmorphic' or 'Not dysmorphic'. All physicians had no specific dysmorphology training beyond their standard medical training and had been training specifically in neurology for at least

three years. Any faces known to the physician were excluded from analysis. Their clinical judgement of dysmorphism was compared to FSD values in the detection of SVs.

One physician (Jan Novy) also reviewed all unprocessed 3D face images in group 2, blinded to all clinical details. He judged if there was any detectable acquired facial deformity that could be seen and where each was located. At the same time, he also assessed if the lips were apart and if there was sufficient coverage of the face for the DSMs, after relevant training. Patients deemed to have acquired facial deformity were then excluded from the relevant DSMs with reference to the regions covered by each one (Figure 20). The regions affected were decided by the evaluating physician for each image. For example, a patient with a suspected previous nasal fracture would typically be excluded from the Face2 and Nose2 DSMs but not from Eyes2.

#### 4.3.4 MOLECULAR ANALYSIS

Patients were only included if they had undergone aCGH as part of research or clinical workup, or had had genome-wide SNP genotyping as part of earlier research studies.

Oligonucleotide aCGH was performed using the Nimblegen 135K Microarray v3.1 (Nimblegen, Madison, WI, USA) or the Agilent 44 K/60 K/75 K/105 K microarrays (Agilent Technologies) in an accredited clinical laboratory in accordance with manufacturer's instructions. Pathogenicity of SVs, using internal and external databases, and confirmatory testing, such as fluorescence in situ hybridization and/or karyotyping, were determined by the laboratory.

Genome-wide genotyping of SNPs was performed using the Illumina Human 610-Quad array (Illumina, San Diego, CA, USA) at the Institute for Genome Sciences and Policy Genotyping Facility, Duke University, USA with quality control measures across all batches. Further analysis for pathogenic SVs was performed using the PennCNV algorithm version 2008 June 26 as described previously (Heinzen et al., 2010).

Standard PennCNV software quality checks were followed with CNV calls excluded based on: log R ratio standard deviation of >0.28, B allele frequency >0.55, B allele frequency <0.45, B allele frequency drift >0.002, PennCNV confidence score <10, CNV based on <10 probes or CNV crossing the centromere. All deletions in known epilepsy hotspots or greater than 1Mb were visually inspected by a geneticist.

If parents had also undergone testing, the laboratory criteria were used to determine if a SV was inherited or arose *de novo*.

#### 4.3.5 STATISTICS

For the primary endpoint of a significant difference in FSD, the data were assessed for normality and then compared using the Mann-Whitney test. For categorical data on ethnicity and intellectual disability, the Kruskal–Wallis test was used. Fisher’s exact test was performed for differences in other categorical data. A receiver operating characteristic curve was calculated to assess sensitivity and specificity of the models. Bonferroni correction was applied for multiple comparisons. Analysis was conducted using SPSS (versions 19-23, SPSS Inc; Chicago, IL, USA).

### 4.4 RESULTS

#### 4.4.1 SUBJECT POPULATION

In group 1, 163 patients with epilepsy were recruited in total, of whom 30 were excluded due to lack of age-matched or ethnically matched controls and 15 due to lack of genotype data (Table 22). Pathogenic SVs were present in 38 patients and this subset was first compared to the remaining 80 patients without any SVs. The patients with SVs were younger, but the use of age-matching with controls accounts for this in all analyses.

	Patients with pathogenic SVs	Patients without pathogenic SVs	Control subjects
Group 1	163 total		N/A
Excluded due to no genotyping; n (%)	15 (9.2%)		-
Excluded due to age; n (%)	2 (1.2%)	22 (13.5%)	-
Excluded due to ethnicity; n (%)	2 (1.2%)	4 (2.4%)	-
Number included; n (%)	38 (23.3%)	80 (49.1%)	-
Additional patients for Group 2	81 total		N/A
Excluded due to age; n (%)	0	11 (14.7%)	-
Excluded due to ethnicity; n (%)	1 (16.7%)	6 (8.0%)	-
Number included; n (%)	5 (83.3%)	58 (77.3%)	-
Total number included	43	138	388
Age; mean age years (range)	25.8 (3.3 - 53.9)	38.8 (2.8 - 56.3)	21.3 (2.4 - 53.2)
Adults >18 years; n (%)	29 (67%)	135 (98%)	207 (53%)
Males; n (%)	23 (53%)	54 (39%)	196 (51%)
MRI findings; n (%)			
Normal	16 (37%)	51 (37%)	
Incidental findings	2 (5%)	8 (6%)	
Abnormal	19 (44%)	75 (54%)	
Not performed/unavailable	6 (14%)	4 (3%)	
Intellectual disability; n (%)			
Normal/mild	22 (51%)	131 (95%)	
Moderate	7 (16%)	4 (3%)	
Severe/profound	12 (28%)	3 (2%)	
Unknown	2 (5%)	0	
Detection method; n (%)			
aCGH	31 (72%)	21 (15%)	-
SNP array	8 (19%)	117 (85%)	-
FISH/karyotyping	4 (9%)	0	
Centre; n (%)			
London	19 (44%)	135 (98%)	388 (100%)
Brussels/Leuven	6 (14%)	-	-
Florence	18 (42%)	3 (2%)	-

**Table 22 Summary of all subjects who were recruited for 3D stereophotogrammetry, the number of subjects excluded and the number of subjects analysed in dense surface models. Group 1 was used to create the first set of DSMs for training. The validation set, Group 2, included all 163 Group 1 patients and 81 additional patients in the DSM model. Only the additional untested patients were used for validation and therefore their characteristics are shown above. In the group with pathogenic structural variants, children were included, there were more male subjects, and subjects were recruited from three different centres. All patients were matched to control subjects based on age and sex for further analyses. MRI findings, intellectual disability and detection methods were obtained from clinical records and investigation reports**

In group 2, there were 244 patients in total, which includes 163 from group 1 and 81 additional patients. Only these additional 81 untested patients comprised the validation set in the comparisons below. Of the 81, 18 were excluded due to lack of matched controls. The remaining 63 patients had all undergone genotyping and comprised 5 patients with known pathogenic SVs and 58 without them.

All of the patients with pathogenic SVs are detailed in Table 23. The variants range in size from 0.13Mb to 48.1Mb. They comprised 28 deletions, 15 duplications, 2 translocations, 2 inversions and 1 triplication.

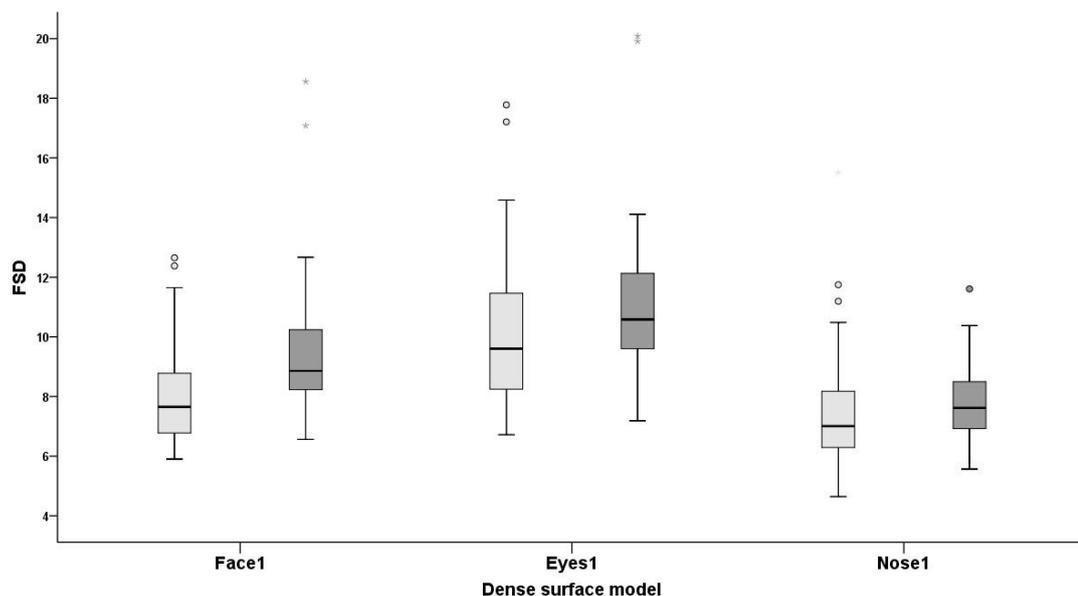
	Sex	Age (yrs)	SV type	Size (Mb)	Origin	Method used	Total genes in SV	FB genes	Face FSD
01	Male	3.3	20pterp13 dup	1.2	de novo	aCGH	25	2	8.2
02	Male	5.8	14q24qter dup	29.5	de novo	aCGH	132	16	9.8
03	Male	7.7	2q22q23 del	10.1	de novo	aCGH	15	6	17.7
04	Male	10.7	11p15q13 inv	N/A	de novo	FISH/karyo	N/A	N/A	11.9
05	Male	10.7	17q21 del 13q14 dup	0.50 0.14	de novo	aCGH	7	1	13.6
06	Male	12.6	20p12 dup	4.9	de novo	aCGH	14	4	8.2
07	Male	19.4	16p13 del	0.81	unknown	SNP array	7	1	12.0
08	Male	22.6	4pterp13 dup 4q35qter del	41.8 6.7	de novo	aCGH	226	33	24.8
09	Male	23.3	2p13 del	0.51	de novo	aCGH	1	0	11.4
10	Male	24.1	15p;19p trans	N/A	unknown	FISH/karyo	N/A	N/A	9.3
11	Male	24.1	15p;19p trans	N/A	unknown	FISH/karyo	N/A	N/A	12.4
12	Male	24.6	2p16 del	0.17	unknown	aCGH	0	0	11.0
13	Male	25.6	16p13 del	0.26	unknown	SNP array	14	1	10.2
14	Male	27.5	15q13 del	1.5	unknown	SNP array	9	0	8.1
15	Male	27.5	14q35 del	3.6	unknown	aCGH	8	1	10.6
16	Male	29.2	17q12 del	2.1	unknown	aCGH	23	0	11.9
17	Male	31.8	1p36 del	2.3	de novo	aCGH	49	3	11.4
18	Male	33.2	1p21 del	5.6	unknown	SNP array	23	2	8.6
19	Male	36.0	15 dup inv	N/A	de novo	FISH/karyo	N/A	N/A	9.8
20	Male	37.3	15q dup	5.6	de novo	aCGH	15	5	11.1
21	Male	42.4	16p12 del	1.1	unknown	SNP array	3	0	10.5
22	Male	44.8	15q11 del 16p13 del	1.2 0.83	unknown	SNP array	14	1	8.9
23	Male	45.8	15q11 del	0.99	unknown	SNP array	5	0	7.7
24	Female	2.1	1q21q23 dup	11.3	de novo	aCGH	71	11	12.4
25	Female	3.5	8p23 dup 17q12 dup	0.35 1.4	inherited	aCGH	18	1	11.1
26	Female	6.9	3p36 dup 17p11 dup	0.83 3.4	unknown	aCGH	51	3	9.2
27	Female	13.6	22q13 dup	6.9	de novo	aCGH	75	8	11.4
28	Female	16.5	21q22 dup	0.13	inherited	aCGH	2	0	20.6
29	Female	17.0	22q11 del	2.5	de novo	aCGH	54	1	11.0
30	Female	17.1	9q22 del	4.4	de novo	aCGH	38	3	10.5
31	Female	17.4	1q44 del	2.6	de novo	aCGH	14	2	14.1
32	Female	18.0	15q26 del	2.23	de novo	aCGH	15	4	7.3
33	Female	22.9	7q22.1 trip	0.13	inherited	aCGH	1	0	10.9
34	Female	24.2	16p13 del	0.74	unknown	aCGH	7	1	10.9
35	Female	24.3	4p dup	48.6	de novo	aCGH	111	12	12.1
36	Female	24.3	3q11 del	4.3	unknown	aCGH	13	0	8.2
37	Female	35.0	11p13p14 del	0.63	unknown	aCGH	5	0	9.2
38	Female	40.9	22q13 del	3.5	de novo	aCGH	42	1	10.6
39	Female	41.0	15q11q13 dup	12.1	unknown	aCGH	49	7	11.9
40	Female	43.8	12p13 del	1.57	unknown	aCGH	13	1	8.1
41	Female	48.5	15q11q13 del	5.4	de novo	aCGH	14	6	13.2
42	Female	49.1	16p13 del	1.5	unknown	SNP array	1	0	9.4
43	Female	53.9	6q22 del	4.1	unknown	aCGH	17	5	12.1

**Table 23 Detailed characteristics of all patients with pathogenic structural variants included in the study. Details of demographics, the type of SV, size in megabases of DNA (Mb), the inheritance or origin of the SV, the number of genes contained in the SV interval, and the whole face FSD value in the Face2 model. The method of detection refers to aCGH, FISH/karyotyping ('FISH/karyo') or genome-wide SNP array ('SNP array'). 'FB genes' refers to the number of genes within the pathogenic SV interval that are known to be highly expressed in the fetal forebrain, according to the Human Brain Transcriptome public database (<http://hbatlas.org/>). All patients were in group 1 (training set), except for the following cases which were only present in group 2 and used for validation: 32, 33, 39, 41**

**and 43. Individuals 8 and 28 were outliers, but FSD was still significant in models for all three face regions after excluding both from analysis.**

#### 4.4.2 FACE SHAPE DIFFERENCE IN PATIENTS WITH AND WITHOUT PATHOGENIC SVS IN THE TRAINING SET

I looked to see if FSD could detect a difference between those with pathogenic SVs and those without. For each of the three DSMs of group 1 (Face1 for the whole face, Eyes1 for the periorbital region, Nose1 for the perinasal region), I calculated FSD for every patient. Those with pathogenic SVs were then compared to those without pathogenic SVs. The median FSD was significantly greater in those with pathogenic SVs than those without, for all DSMs (Figure 39; whole face: 8.86 vs 7.65;  $p=0.001$ , periorbital region: 10.6 vs 9.60;  $p=0.013$ , perinasal region: 7.62 vs 7.01;  $p=0.031$ , for pathogenic SV vs no pathogenic SV respectively).



**Figure 39** Box plots of the median, interquartile range and range of Face Shape Difference (FSD) for the three different models using the training cohort ( $n=118$ ). FSD is significantly greater for the whole face model (Face1; 8.86 vs 7.65,  $p=0.001$ ), the periorbital model (Eyes1; 10.6 vs 9.60,  $p=0.013$ ), and the perinasal model (Nose1; 7.62 vs 7.01,  $p=0.031$ ) in patients with pathogenic SVs. Outliers over 1.5 or 3 times the interquartile range from the upper quartile are shown in circles or asterisks, respectively. Excluding all outliers does not alter significance.

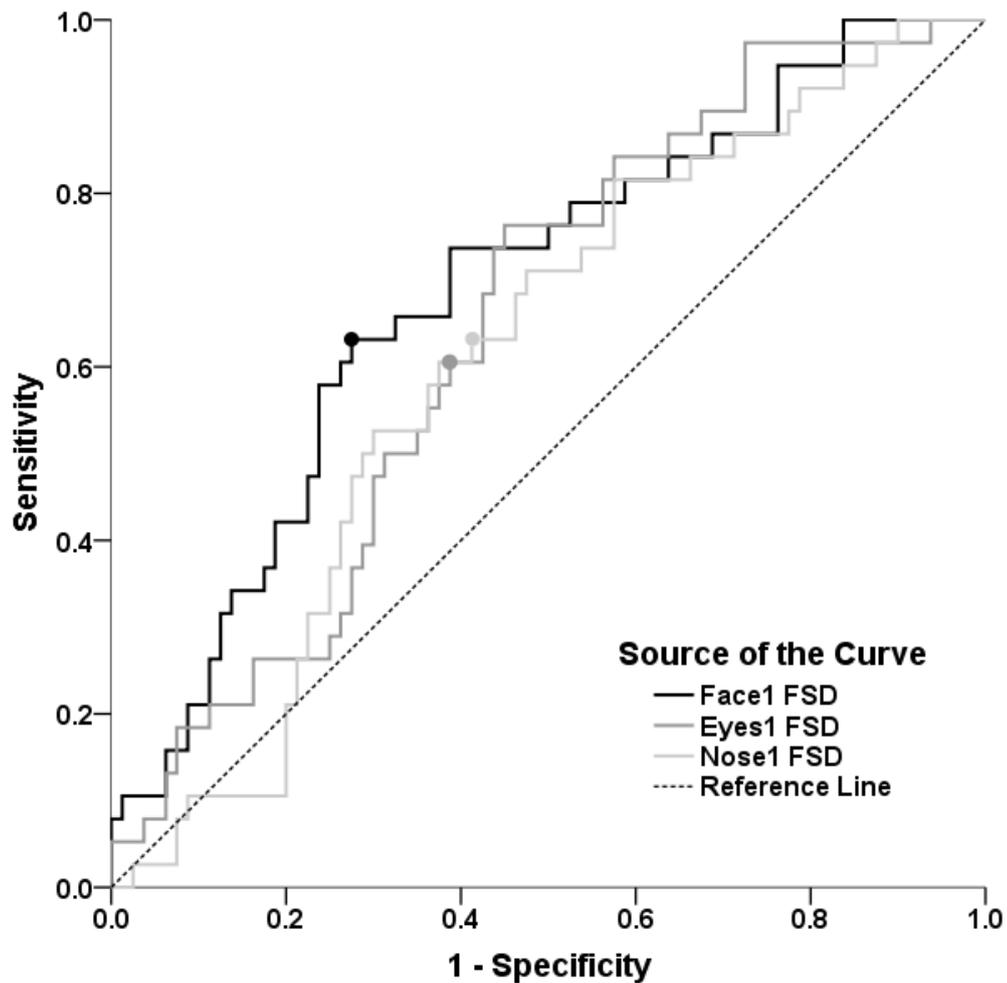
The distribution of FSD values shows outliers for all models, in groups with and without pathogenic SVs. After exclusion of all outliers, FSD was still significantly greater in those with pathogenic SVs (whole face:  $p=0.001$ , periorbital region:  $p=0.018$ , perinasal region:  $p=0.018$ ). Another sensitivity analysis excluding outliers in the combined training and validation set (group 2) is discussed in Section 4.4.6.

FSD of the whole face showed a strong positive correlation with the periorbital region ( $\rho=0.78$ ;  $p<0.001$ ). The perinasal region is less strongly correlated with FSD in other facial regions ( $\rho=0.50$ ;  $p<0.001$  with whole face,  $\rho=0.60$ ;  $p<0.001$  with periorbital region). This suggests that FSD of the whole face, periorbital region or perinasal region is significantly greater in people with epilepsy and pathogenic SVs, than in people with epilepsy without such variants.

#### 4.4.3 FACE SHAPE DIFFERENCE IN PATIENTS WITH AND WITHOUT PATHOGENIC SVS IN THE VALIDATION SET

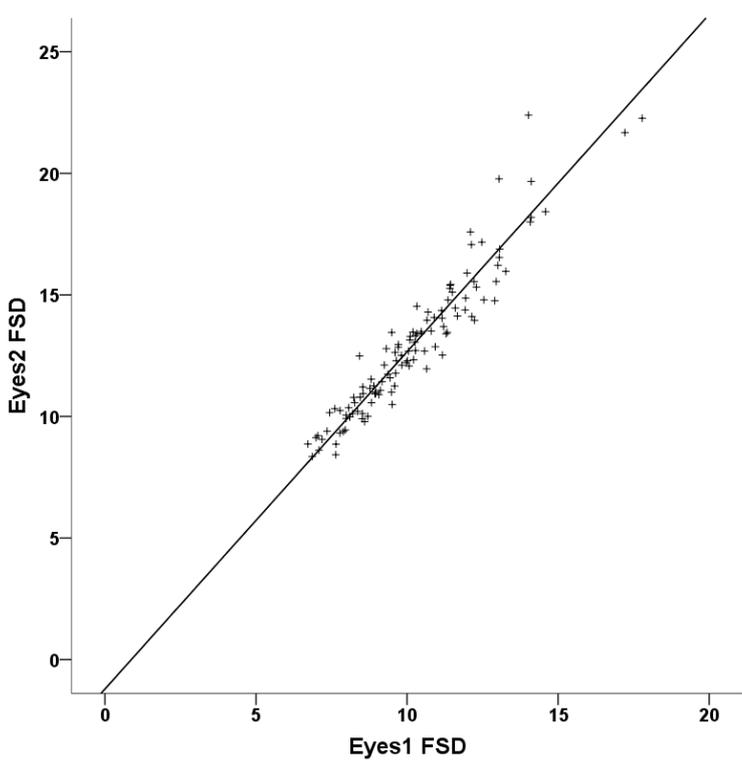
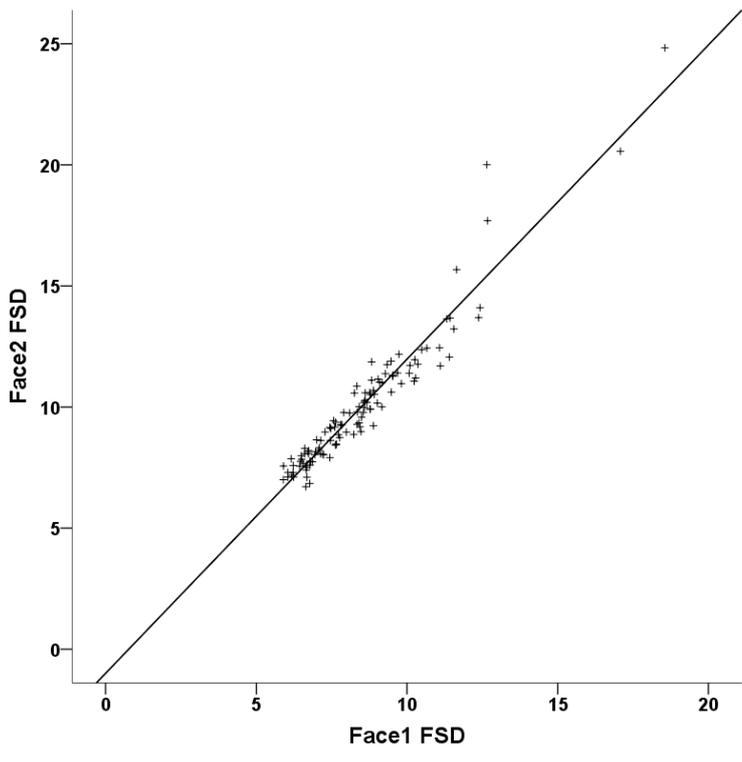
Next, I determined an optimum FSD cut-off value for detection of SVs using the training set from group 1 and judged accuracy of FSD for a new validation set. From the individual FSD values for those with and without pathogenic SVs, receiver operating characteristic (ROC) curves were created to see how accurate FSD is for differentiating those with SVs (Figure 40). The area under the curve was 0.69 (95% CI 0.60-0.80;  $p<0.001$ ) for the Face1 model. An FSD value of 8.47 was found to be the optimal cut-off value with equal sensitivity and specificity (65.8%) for determining if an individual face image is from a patient with a pathogenic SV. FSD cut-off values were found for the Eyes1 and Nose1 models using the same method (

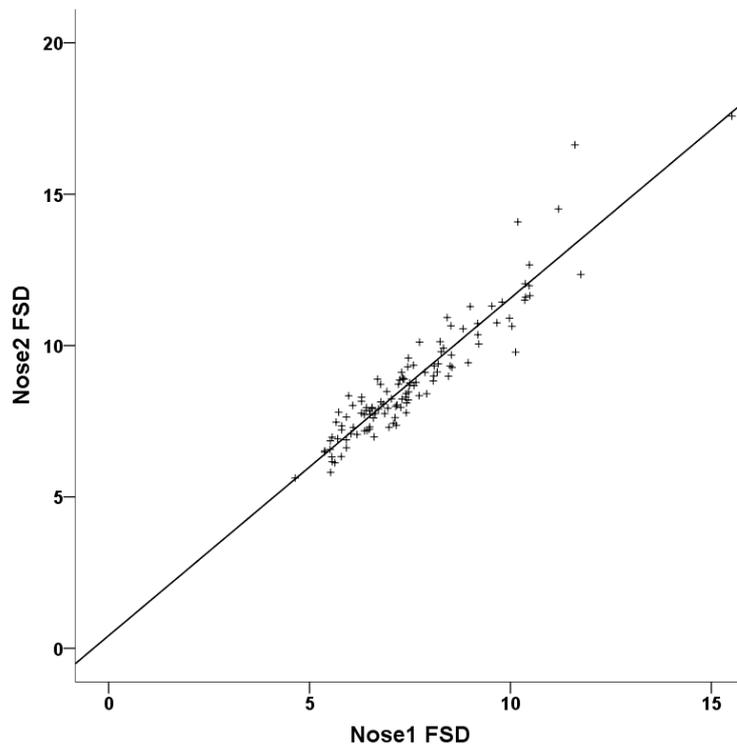
Table 24). As stated earlier, FSD values are specific to individual DSMs due to differing principal components (PCs) between DSMs. Therefore, I had to identify the equivalent FSD for the group 2 DSMs before validation could be tested in this group.



**Figure 40 Receiver operating characteristic (ROC) curves of face FSD, periorbital FSD, and perinasal FSD used for detecting pathogenic SVs in the training cohort. The areas under the curve are 0.69, 0.64, and 0.61 respectively. The filled circles mark the optimal FSD cut-off values for equal sensitivity and specificity. These cut-off values were used for prediction in the validation set within group 2 (Face2, Eyes2, Nose2 DSMs)**

For the 118 patients who were each in group 1 and in group 2, FSD values in the original and second set of models showed very strong positive correlation as expected (Figure 41; Face1 vs Face2:  $\rho=0.96$ ;  $p<0.001$ , Eyes1 vs Eyes2:  $\rho=0.96$ ;  $p<0.001$ , Nose1 vs Nose2:  $\rho=0.93$ ;  $p<0.001$ ;  $n=118$  for all). The relationship between FSD values showed a strong positive correlation, suggesting that these values are highly reproducible. This is in agreement with a reproducibility study of FSD in different DSMs in Section 3.4.3.





**Figure 41** For the 118 patients common to group 1 and group 2, there was very strongly positive correlation for FSD values between the Face1 and Face2 models ( $\rho=0.96$ ;  $p<0.001$ ). This was also true for the periorbital and perinasal region (Eyes1 vs Eyes2:  $\rho=0.96$ ;  $p<0.001$ , Nose1 vs Nose2:  $\rho=0.93$ ;  $p<0.001$ ). Best-fit linear regression lines (shown) were used to convert the optimal FSD cut-off values from the group 1 DSMs to the corresponding group 2 DSMs so that it could be used to predict the presence of pathogenic SVs in the validation cohort. For example, the formula for the line in the first graph is:  $\text{Face2 FSD} = (1.30 \times \text{Face1 FSD}) - 0.99$ .

In the validation set, comprising only the 63 additional untested group 2 patients, the adjusted whole face FSD cut-off value (FSD=9.99) correctly showed that 4 of 63 patients had pathogenic SVs. One additional patient with a pathogenic SV was not detected (i.e. 4/5, 80% sensitivity). Whole face FSD correctly classified 45 patients who were deemed to have no pathogenic SVs (45/58, 78% specificity). Using periorbital and perinasal FSD resulted in reduced sensitivity (3/5, 60% for both) but similar specificity (40/58, 69% and 45/58, 78%, respectively). Results are shown in Table 24.

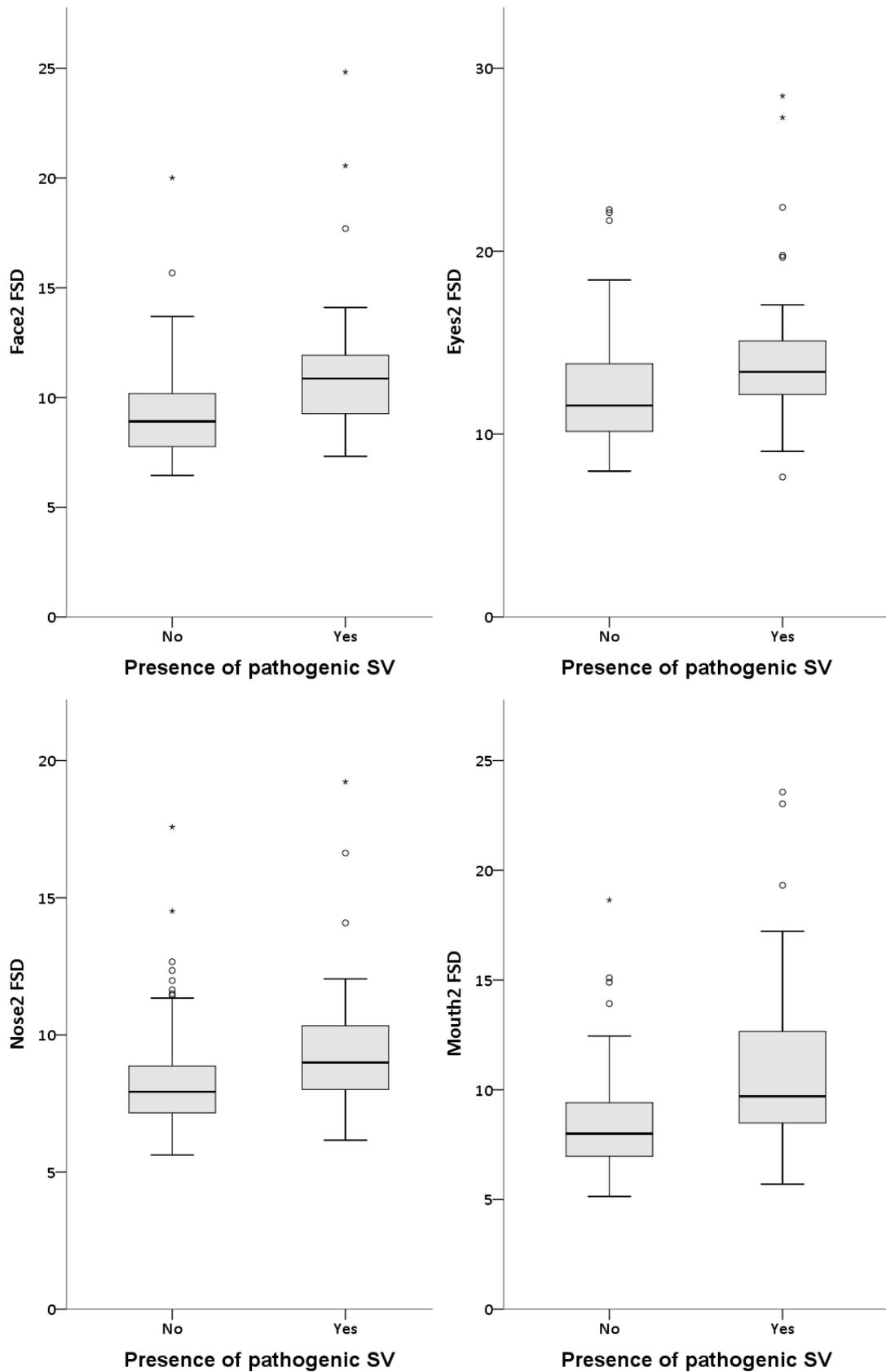
The validation set shows that FSD can discriminate between people with epilepsy with and without SVs with a sensitivity of 60-80% and specificity of 69-78%, depending on the region of the face analysed.

<b>Face region</b>	Whole face	Periorbital region	Perinasal region
<b>Group 1 DSMs</b>	<b>Face1</b>	<b>Eyes1</b>	<b>Nose1</b>
Number in training set	118	118	118
Area under the curve	0.69	0.64	0.61
Cut-off value	8.47	10.22	7.39
Predicted sensitivity	66% (25/38)	61% (23/38)	63% (24/38)
Predicted specificity	65% (52/80)	61% (49/80)	63% (50/80)
<b>Group 2 DSMs</b>	<b>Face2</b>	<b>Eyes2</b>	<b>Nose2</b>
Number analysed in validation set	63	63	63
Equivalent cut-off value	9.99	12.96	8.66
Actual sensitivity	80% (4/5)	60% (3/5)	60% (3/5)
Actual specificity	78% (45/58)	69% (40/58)	78% (45/58)
Positive predictive value	26% (4/17)	14% (3/21)	19% (3/16)
Negative predictive value	98% (45/46)	95% (40/42)	96% (45/47)

***Table 24 The table shows all 6 DSMs of the 3 different facial regions. The top half is for DSMs of group 1, the training set, which were used to calculate a FSD cut-off value with equal sensitivity and specificity. This FSD cut-off value was translated into an equivalent FSD value in the group 2 DSMs using the 118 patients present in both groups. This translation is necessary because FSD values are only specific for the DSM they are calculated from, due to unique principal components for each DSM. This is further explained in Chapter 3. Then this adjusted FSD cut-off for group 2 DSMs was used to predict the presence or absence of pathogenic SVs in the 63 additional untested patients. Prediction was most accurate using the whole face DSM (Face2) with sensitivity of 80% and specificity of 78%. The periorbital and perinasal regions are both less sensitive and less specific.***

Group 2 DSMs were reassessed using all 181 patients: 43 with pathogenic SVs and 138 without – to see if FSD was still significantly different in those with SVs as it was in group 1. On this occasion, due to developments in the DSME Explorer software program, it was possible to use a new DSM of only the perioral region. FSD was significantly

greater in patients with pathogenic SVs than in those without for all four DSMs of the face (whole face: 10.9 vs 8.91;  $p < 0.001$ , periorbital region: 13.4 vs 11.6;  $p < 0.001$ , perinasal region: 8.99 vs 7.93;  $p < 0.001$ , perioral region: 9.70 vs 8.00;  $p < 0.001$ , for pathogenic SV vs no pathogenic SV respectively). This is shown in Figure 42.



**Figure 42** Box plots of FSD values in all patients in group 2 according to the presence of pathogenic SVs, in four DSMs of the whole face (Face2), periorbital region (Eyes2),

*perinasal region (Nose2) and perioral region (Mouth2). FSD was significantly greater in those with pathogenic SVs for all four DSMs.*

#### 4.4.4 FACE SHAPE CHARACTERISTICS IN PEOPLE WITH PATHOGENIC STRUCTURAL VARIANTS

Given that FSD was increased, I analysed whether any of the people with pathogenic SVs had any common face shape characteristics. Of the 43 people recruited with pathogenic structural variants and epilepsy, 31 had pathogenic SVs unique to them within this group. Previous studies of people with epilepsy have also found that many people with epilepsy have individual or unique structural variants within the sample (Helbig et al., 2014). No shared facial features were expected given the diversity of SVs. Nevertheless, DSMs may identify subtle facial features common to all those with SVs. Therefore, each DSM had each PC (81 in Face2, 172 in Eyes2, 70 in Nose2) tested for any significant difference between patients with pathogenic SVs and those without. Bonferroni correction was applied for multiple testing. No individual PC was significantly different in those with pathogenic SVs and those without, in any of the models.

The most common recurring pathogenic SV in the group was for 16p13.11 deletion in five cases. For only these five cases, there was no significant difference in FSD of the whole face compared to patients with no pathogenic SVs (10.2 vs 8.91; Mann-Whitney U test;  $p=0.09$ ;  $n=143$ ). There was also no significant association with any individual PC for the five patients.

A composite average face can be created for a group of people by using the mean of every principal component (PC) for face images in that group. The average face can then be viewed in the DSMExplorer software. Upon inspection of the average face of patients with pathogenic SVs, there was no obvious facial dysmorphism. This was further analysed by colour-coding the average face for the difference in distance of each

surface point from the average face of people without pathogenic SVs. Therefore, there was no evidence for shared facial features in patients with pathogenic SVs.

#### 4.4.5 COMPARISON OF FSD WITH INSPECTION OF FACES BY NEUROLOGISTS

Four physicians assessed all patients in group 1, except for one physician (Jan Novy) who was unable to assess the images of five patients because they were known to him. Their results are shown in Table 25. Sensitivity and specificity of their labelling of face images as 'Dysmorphic' were 47-74% and 74-94%, respectively.

	Judgement	Physician 1	Physician 2	Physician 3	Physician 4
Patients with SVs	Dysmorphic	28	25	21	17
	Not dysmorphic	10	13	17	19
Patients without SVs	Dysmorphic	21	19	18	5
	Not dysmorphic	59	61	62	72
	Sensitivity	74%	66%	55%	47%
	Specificity	74%	76%	78%	94%
	Positive predictive value	57%	57%	54%	77%
	Negative predictive value	86%	82%	78%	79%

**Table 25 Results from categorisation of all patients in group 1 as either 'Dysmorphic' or 'Not dysmorphic' by four adult neurologists in training. Sensitivity ranges from 0.46 to 0.73 and specificity ranges from 0.74 to 0.94. Neurologists were blinded to all clinical details and assessed 3D face images only. They were also told that there was at least one person with known facial dysmorphism in the group.**

#### 4.4.6 EXCLUSION OF ALL FSD VALUES THAT ARE OUTLIERS

For patients in group 2, a number of different factors were assessed in sensitivity analyses. Firstly, there are extreme outliers in all of the 4 DSMs that can be seen on the box plots (Figure 42). These outliers are defined using standard criteria as those values which lie more than 1.5 times the interquartile range above the upper quartile or 1.5 times the interquartile range below the lower quartile. Outliers with and without

pathogenic SVs may bias the results and increase the chances of a type 1 or type 2 error, and so a repeat analysis was done excluding all outliers. Upon exclusion of all outliers for each DSM, FSD was still significantly increased in patients with pathogenic SVs (Table 28).

#### 4.4.7 EXCLUSION OF IMAGES FROM CHILDREN AND EUROPEAN SITES

Next I controlled for possible bias from ethnic differences or the higher number of children in the group with pathogenic SVs. I excluded all face images from people recruited outside the UK. This comprised 21 subjects from Italy and 6 from Belgium. Only the European sites had recruited children into the study. Therefore, by excluding these sites, the sensitivity analysis was restricted to white British adults only. FSD was significantly increased for all four DSMs (Table 26). Thus, ethnicity differences at the different recruitment sites did not appear to be a confounder for the results for FSD.

	<b>FSD in UK adults with pSVs (n)</b>	<b>FSD in UK adults without pSVs (n)</b>	<b>Model</b>	<b>Significance</b>
Whole face	11.0 (19)	8.87 (135)	Face2	p =0.001*
Periorbital region	13.4 (19)	11.7 (135)	Eyes2	p =0.007*
Perinasal region	9.29 (19)	7.92 (135)	Nose2	p =0.002*
Perioral region	9.70 (19)	7.99 (135)	Mouth 2	p =0.004*

**Table 26 FSD values in white British adults with and without pathogenic SVs (pSVs) for all four DSM models of the face. FSD was significantly increased in all DSMs in patients with pathogenic SVs.**

#### 4.4.8 EXCLUSION OF IMAGES WITH FACIAL INJURY

Overall 37 images were excluded on the basis of visual inspection by a physician and these comprised 46 acquired deformities, with 7 images thought to have two categories of deformities and 1 image thought to have three categories, shown below (Table 27). No images were found to have inadequate face coverage or excessive facial hair. All DSMs containing the injury in question were then excluded. For example, in a patient

with a scar over their eyebrows, the whole face DSM and periorbital DSM were excluded but not the perinasal and perioral DSMs, which would not be expected to have altered. Analysis of FSD in the remaining images still showed a significant difference between patients with pathogenic SVs and those without (Table 28). Facial injury did not appear to be a confounder in this study.

Acquired deformities	Number	Notes
Periorbital scar	14	1 pierced eyebrow; rest on eyebrow or eyelid
Upper face scar	10	Usually forehead
Nose fracture	7	
Lower face scar	4	Usually cheek or chin scars
Nasal scar	3	
Perioral scar	3	
Face asymmetry	2	1 facial palsy, 1 with right temporalis wasting
Skin lesions	2	Both raised lesions
Contusion	1	On forehead

**Table 27 Categories of acquired facial deformities seen in all patients in group 2.**

#### 4.4.9 EXCLUSION OF IMAGES WITH MOUTH OPEN

Nineteen patients had their mouth open during image capture. In the previous chapter, it was seen that facial expression could influence landmark placement and also FSD. After excluding all of these patients, FSD remained significantly different between the patients with pathogenic SVs and those without in all DSMs (whole face:  $p=0.002$ , periorbital region:  $p=0.011$ , perinasal region:  $p=0.001$ , perioral region:  $p=0.004$ ).

Another analysis of only those who had their mouth closed and had no acquired facial deformity was conducted. In this analysis, anyone with any acquired deformity was excluded from all DSMs, not just the ones thought to be affected. This was because my earlier reproducibility study in Chapter 3 had suggested that FSD of regions of the face could be affected by facial expressions elsewhere – such as the periorbital FSD by mouth opening, despite unchanged periorbital landmark distances. This analysis also showed that FSD remains significantly increased in the presence of pathogenic SVs, in people with epilepsy (Table 28).

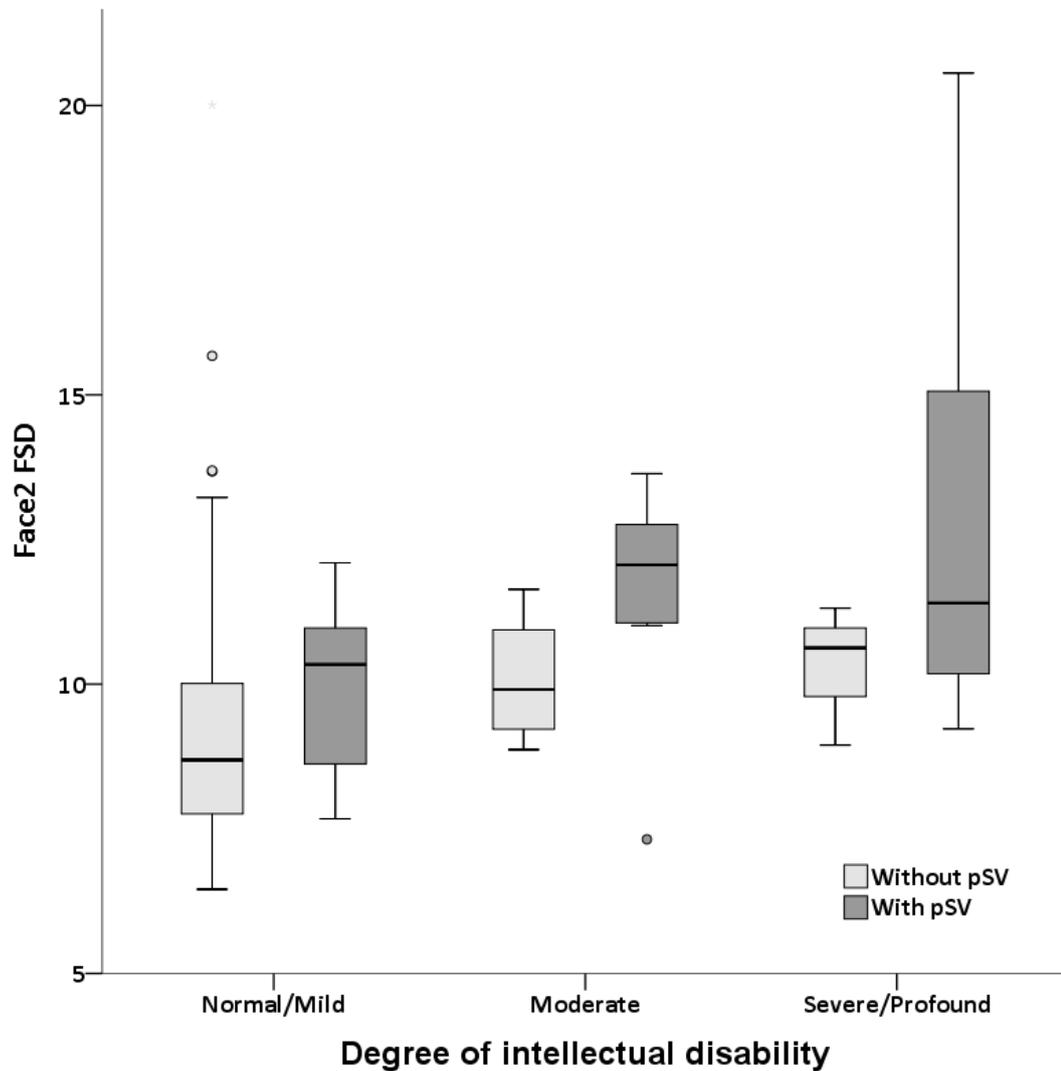
#### 4.4.10 EXCLUSION OF IMAGES IN PATIENTS WITH MODERATE-SEVERE INTELLECTUAL DISABILITY

Intellectual disability may also be a confounder in this study. People with moderate or severe intellectual disability may be less likely to be able to adopt a neutral head position, to be able to keep a neutral facial expression, or to understand instructions for image capture, such as to keep their eyes open. Whilst these factors were all taken into consideration at the time of image capture, people may have had less noticeable changes in head position or posture or face expression. Head position was found to affect FSD in Chapter 3. Only those with an objective change in facial expression (i.e. with lips apart) were excluded in the previous section, and people who may be smiling, smirking, pouting or any other non-neutral expression with their lips together may not have been excluded.

Data on intellectual function was available for 179 out of 181 subjects. For 127, this was derived from formal neuropsychometry assessment and for another 52, it was from clinical records alone. Twelve people with moderate or severe/profound intellectual disability were in the group with pathogenic SVs whereas only three people had moderate to profound intellectual disability without a pathogenic SV. Exclusion of those with greater than mild intellectual disability did not change the significance of FSD findings for any DSM (whole face:  $p=0.007$ , periorbital region:  $p=0.047$ , perinasal region:  $p=0.004$ , perioral region:  $p=0.041$ ). The sensitivity analysis is shown in Table 28.

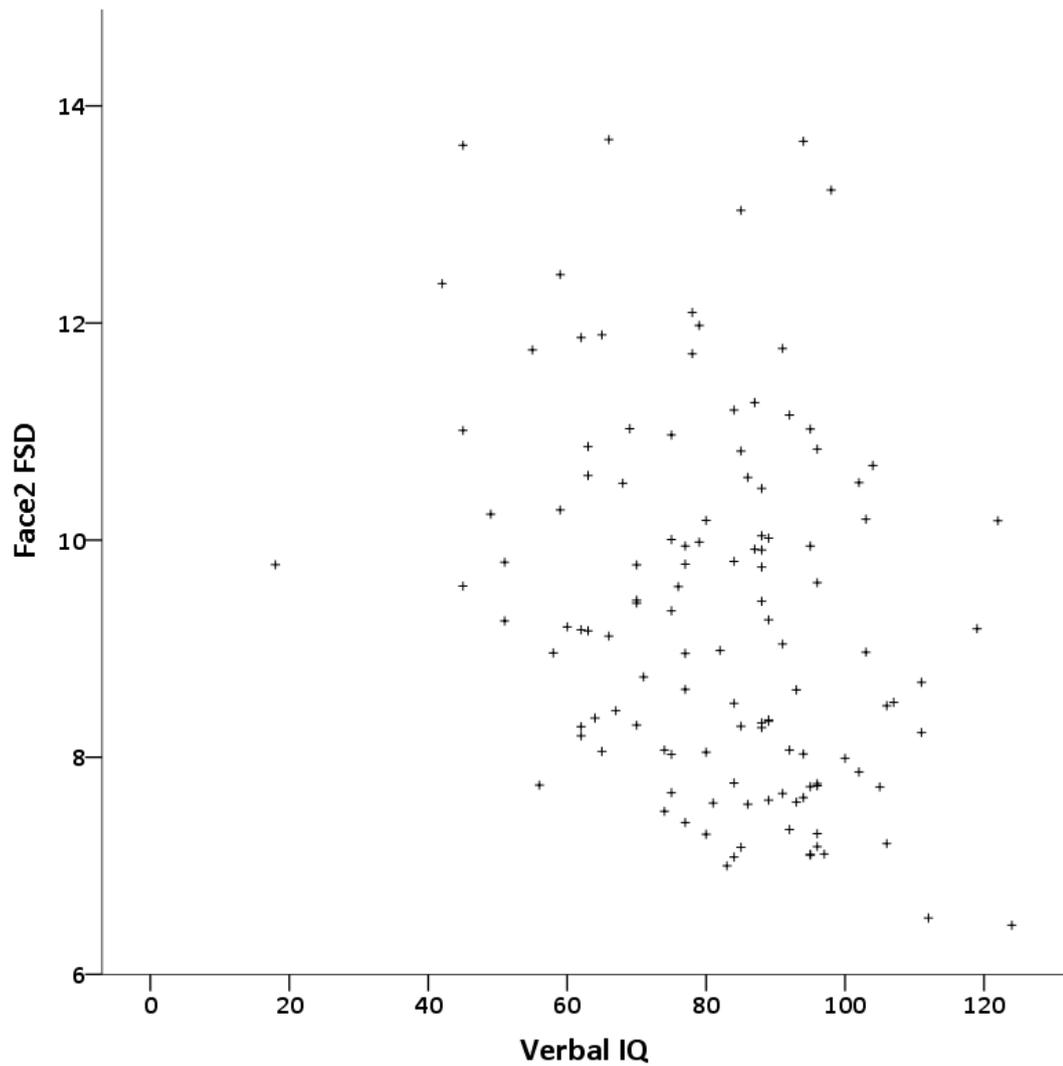
Intellectual disability was classified in my study population on the basis of formal neuropsychometry assessment in 127 subjects, and clinical records alone in 52 further individuals. Two patients could not be categorised. FSD in the Face2 DSM for the whole face was significantly different for the three categories of intellectual disability (8.96 vs

11.1 vs 11.3; independent samples Kruskal-Wallis test;  $p < 0.001$ , for normal/mild ID, moderate ID, severe/profound ID, respectively). FSD also differed significantly for the periorbital, perinasal and perioral regions alone. FSD appears to only increase with the severity of intellectual disability in the group of people with pathogenic SVs, but this is not true of those without (Figure 43).



**Figure 43** Box plots showing the association between FSD and intellectual disability.

For those who underwent formal neuropsychometric estimation of IQ, there was also a significant but weakly negative correlation between IQ score and FSD ( $\rho = -0.31$ ;  $p = 0.001$ ;  $n = 122$ ; shown in Figure 44).



**Figure 44** Scatter plot of verbal IQ as measured on formal neuropsychometry testing with Face2 FSD in 127 patients for whom data was available. There is a weak negative correlation between IQ score and FSD ( $\rho=-0.31$ ).

	<b>FSD in patients with pSVs (n)</b>	<b>FSD in patients without pSVs (n)</b>	<b>Model</b>	<b>Significance</b>
<b>Exclusion of outliers</b>				
Whole face	10.7 (40)	8.82 (136)	Face2	p<0.001
Periorbital region	13.1 (38)	11.4 (134)	Eyes2	p=0.002
Perinasal region	8.91 (40)	7.87 (130)	Nose2	p<0.001
Perioral region	9.64 (40)	7.98 (134)	Mouth2	p<0.001
<b>Physician review for deformity</b>				
No face insult	11.0 (37)	8.64 (107)	Face2	p <0.001*
No periorbital insult	13.2 (40)	11.2 (126)	Eyes2	p <0.001*
No perinasal insult	8.99 (43)	7.87 (128)	Nose2	p <0.001*
No perioral deformity	9.94 (41)	8.00 (135)	Mouth2	p <0.001*
<b>No face deformity and lips closed</b>				
Whole face	10.6 (25)	8.63 (102)	Face2	p =0.001*
Periorbital region	13.1 (25)	11.0 (102)	Eyes2	p =0.012*
Perinasal region	8.94 (25)	7.87 (102)	Nose2	p =0.002*
Perioral region	9.31 (25)	7.68 (102)	Mouth2	p =0.003*
<b>Only normal-mild intellectual disability</b>				
Whole face	10.3 (24)	8.69 (131)	Face2	p =0.007*
Periorbital region	12.8 (24)	11.4 (131)	Eyes2	p =0.047*
Perinasal region	8.86 (24)	7.92 (131)	Nose2	p =0.004*
Perioral region	8.94 (24)	7.95 (131)	Mouth2	p =0.041*

**Table 28 Summary of sensitivity analyses looking at FSD values between patients with pathogenic SVs (pSVs) and those without, after exclusion of various potential confounders. FSD remains significantly increased in patients with pathogenic SVs in all subgroups.**

#### 4.4.11 ANTI-EPILEPTIC DRUG HISTORY

Anti-epileptic drug use may have contributed to part of the difference seen in FSD.

Previous drug history was available for 170 of 181 patients and 158 were adults. Only

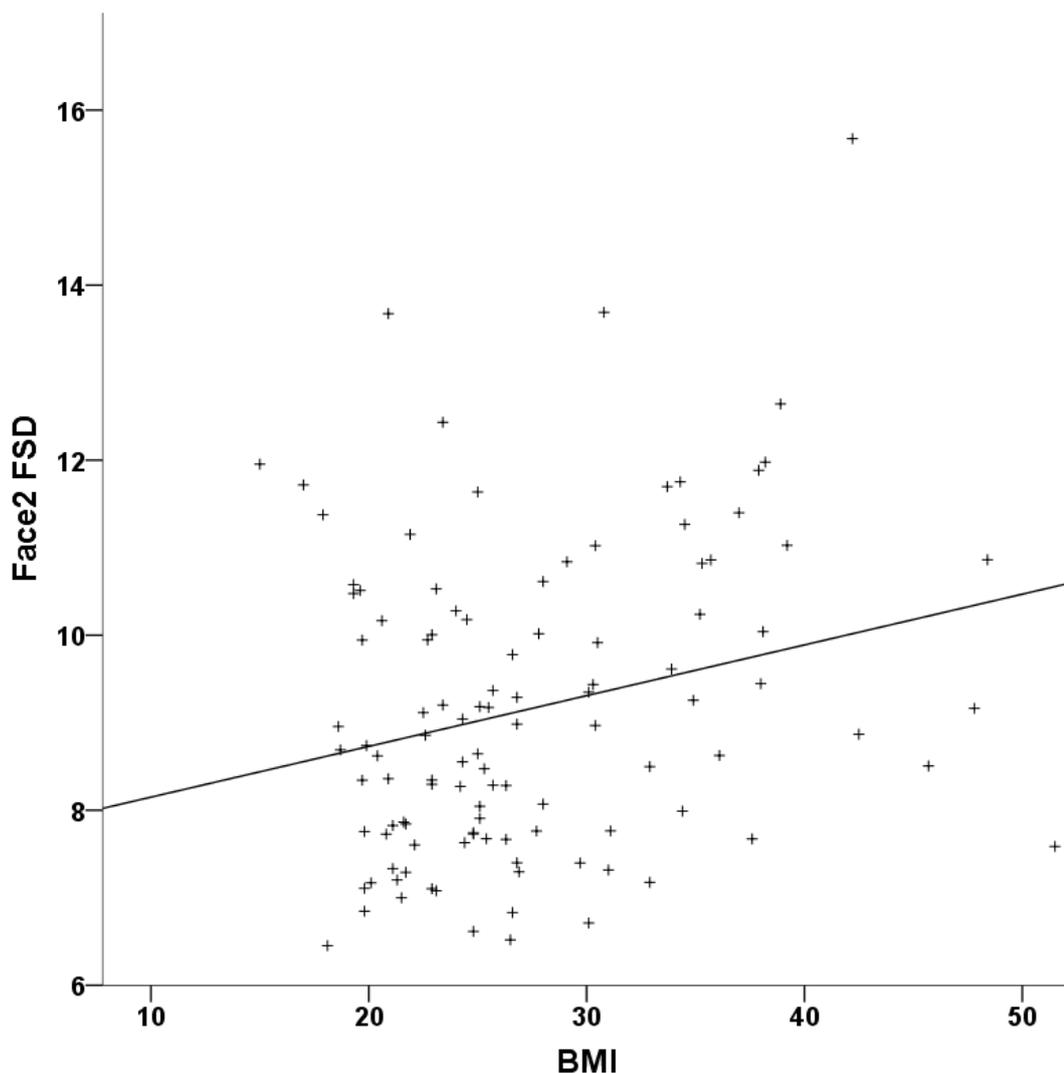
adults were compared to avoid bias from children, for whom different drugs may have been prescribed and there may be fewer previous drugs used. There was no significant difference in the total number of anti-epileptic drugs used between those with pathogenic SVs and those without (median 6 vs 6; n=158; Mann-Whitney U test;  $p=0.12$ ). There was also no significant difference in the proportion of both groups who had used any individual anti-epileptic drug (Table 29).

	Adults with pathogenic SVs	Adults without pathogenic SVs	Significance before correction
Number in group 2	29	135	-
Number with known drug history	27	131	-
Mean age (years)	32.7	40.1	-
Mean no of drugs tried	5.5	6.8	-
Median no of drugs tried	6	6	$p=0.12$
Acetazolamide	4%	24%	$p=0.02$
Carbamazepine	67%	82%	$p=0.07$
Clobazam	52%	58%	$p=0.67$
Clonazepam	22%	18%	$p=0.59$
Gabapentin	11%	29%	$p=0.06$
Lamotrigine	59%	76%	$p=0.10$
Levetiracetam	59%	81%	$p=0.02$
Oxcarbazepine	15%	34%	$p=0.07$
Phenobarbital	37%	30%	$p=0.50$
Pregabalin	7%	22%	$p=0.11$
Phenytoin	37%	51%	$p=0.21$
Primidone	4%	12%	$p=0.31$
Tiagabine	7%	11%	$p=0.74$
Topiramate	44%	62%	$p=0.13$
Vigabatrin	7%	31%	$p=0.01$
Valproate	81%	73%	$p=0.47$
Zonisamide	15%	16%	$p=1.00$

**Table 29 Comparison of previous or current anti-epileptic drug use between adults in group 2 who had pathogenic SVs and those who did not. Anti-epileptic drug use was not significantly different for the total number of drugs used or for any individual anti-epileptic drug (after correcting for multiple testing). Note that rates of use for phenytoin and sodium valproate in particular are similar in both groups.**

#### 4.4.12 BODY MASS INDEX (BMI)

Information on height and weight was available for 110 out of 181 patients, of whom 15 had pathogenic SVs. There was a weakly positive correlation between FSD in the whole face and BMI (Face2:  $\rho=0.21$ ;  $p=0.027$ ;  $n=110$ ; Figure 45). This was also true for perioral DSM FSD and BMI (Mouth2:  $\rho=0.24$ ;  $p=0.011$ ;  $n=110$ ), but there was no correlation with BMI for FSD in the periorbital and perinasal DSMs.



**Figure 45** Scatter plot of FSD for the whole face DSM and BMI in 110 patients in group 2 for whom height and weight data was available. There is a weakly positive correlation of FSD with BMI shown by the regression line above ( $\rho=0.21$ ;  $p=0.027$ ).

The mean BMI did not vary significantly between subjects with pathogenic SVs and subjects without (28.4 vs 27.2; independent samples t-test;  $p=0.06$ ;  $n=110$ ). As an

additional check, 64 patients who were outside of the normal BMI range (18.5-24.9) were excluded in a sensitivity analysis (Table 30). FSD was still significantly increased in patients with pathogenic SVs for analysis using the whole face, periorbital region or perioral region. FSD of the perinasal region was not significantly different between those with and without pathogenic SVs.

<b>Patients with normal BMI only</b>				
	<b>FSD in patients with pSVs (n)</b>	<b>FSD in patients without pSVs (n)</b>	<b>Model</b>	<b>Significance</b>
Whole face	10.3 (4)	8.32 (42)	Face2	p =0.021*
Periorbital region	14.1 (4)	10.7 (42)	Eyes2	p =0.023*
Perinasal region	9.11 (4)	7.67 (42)	Nose2	p =0.051
Perioral region	10.1 (4)	7.44 (42)	Mouth2	p =0.042*

**Table 30 Sensitivity analysis of all patients in Group 2 with available BMI data who were within the normal BMI range and categorised according to the presence or absence of a pathogenic SV (pSV). There was a significantly increased FSD in all DSMs except the perinasal region in those with pSVs.**

#### 4.5 DISCUSSION

The primary hypothesis that DSMs, and FSD specifically is able to discriminate between people with epilepsy and pathogenic SVs compared to people with epilepsy without pathogenic SVs, was true in this sample for the two groups overall. From my literature review, this appears to be the first time a computer-based analysis of face images has been able to discriminate people with a group of different genetic disorders in conjunction with epilepsy: the group of 43 people with pathogenic SVs had 31 unique and non-overlapping SVs amongst them. FSD was not able to correctly identify all individuals though, with a maximum sensitivity of 80% and maximum specificity of 78%.

Not surprisingly, using the whole face to calculate FSD proved to be the most accurate method with sensitivity of 66-80% and specificity of 65-78%. Regional DSMs using only

parts of the face performed worse with sensitivity of 60-63% and specificity of 61-78% for the periorbital and perinasal DSMs. The decreased sensitivity of the periorbital and perinasal DSMs is in keeping with a previous study comparing them to the whole face DSM using the same software (Hammond et al., 2005). Using the perioral region also seems to be able to differentiate those with SVs. The low positive predictive values (14-26%) and high negative predictive values (95-98%) in all DSMs may suggest the technique is best used to screen for those who may have pathogenic SVs rather than as a diagnostic test.

Previously, DSM was able to distinguish two particular genetic microdeletion syndromes associated with seizures, Smith-Magenis syndrome and Wolf-Hirschhorn syndrome, from controls (Hammond et al., 2012a, 2005). The current findings suggest that DSM may be useful in identifying a much wider range of pathogenic SVs. The most frequent recurring SV was 16p13.11 deletion in five cases and no evidence of atypical facial features was found for just this group in isolation, although there may be a trend towards a higher FSD. Given that 16p13.11 deletion is also a recurrent and relatively common microdeletion in people with epilepsy, and another such microdeletion at 15q13.3 appears to have mild facial dysmorphism, it would be useful to study this deletion in a larger sample (Sisodiya and Mefford, 2011).

FSD is a quantitative objective measurement that could identify novel patterns of abnormality of facial anatomy. DSM and PC analysis, upon which FSD is based, has already been shown to do this (Tobin et al., 2008). This may be because an increase in FSD does not directly translate to clinically observable dysmorphism. FSD is strictly based on face shape analysis. Clinical dysmorphism is based on subjective identification of individual features or a facial 'gestalt' (Hennekam et al., 2010). Recognition of faces by humans appears to be dependent on colour, pigmentation, motion as well as contour, and humans find it difficult to identify faces based on contour alone (Sinha et

al., 2006). In keeping with this, some study participants with high FSD values were not thought to be dysmorphic when evaluated subjectively by their regular physician. I also found no evidence for shared facial features in people with epilepsy associated with diverse underlying genomic abnormalities, which was supported by visual inspection of the average face.

Compared to FSD, four physicians training in neurology could identify patients with pathogenic SVs with a sensitivity of 47-74% and specificity of 74-94%. This is broadly similar to FSD, which may have slightly higher sensitivity and slightly lower specificity. Physicians were told that there were patients with dysmorphism in the group and were specifically primed to look for facial dysmorphism. The freely rotatable 3D face image offers advantages and disadvantages to assessing a real patient, but overall the lack of expression and motion may make it more difficult for physicians.

Evaluation for dysmorphism should be part of the clinical examination in epilepsy. Indeed, one of the indications for aCGH in people with epilepsy is the presence of dysmorphic features (Kharbanda et al., 2015). However, dysmorphism may be missed by untrained clinicians and even clinical geneticists may take years to learn to recognise some patterns of facial dysmorphism (Reardon and Donnai, 2007). Outside of clinical genetics and especially in adult medicine, there is lack of confidence regarding genetic testing and a lack of awareness of dysmorphic features and genomic causes (Burke et al., 2006; Mainous et al., 2013; Maves et al., 2007; Salm et al., 2014; Williams, 2007; Wolyniak et al., 2015). Another factor is that many syndromes have only been recently described and so there may be many adult patients with undiagnosed genomic disorders, including pathogenic SVs, presenting to adult medicine clinics. This has been noted for adult epilepsy (Dixit and Suri, 2016).

A number of potential confounders were also assessed. Evidence of acquired facial deformity was present in 37 patients. All but five of these were scars, contusions or

abnormal nose shape, due to suspected trauma and injury. The remainder included a piercing, skin lesions, and facial asymmetry from suspected Bell's palsy or surgery. The frequency is consistent with the higher prevalence of trauma in people with epilepsy (van den Broek and Beghi, 2004; Wirrell, 2006), although there are two differences. This sample was selected from tertiary epilepsy centres and may have an increased frequency of seizures compared to the general population of people with epilepsy. Secondly, only visual evidence of deformity was assessed by a physician blinded to clinical details, unlike previous studies which looked at clinical history.

Facial expression was associated with differences in FSD in the earlier study of reproducibility. This was most marked for mouth opening. In this study, no patient was smiling or puckering their lips, but a number had their mouths open. This was especially apparent in children and those with intellectual disability, and both of these groups were more frequent in the group with pathogenic SVs. Thus, it was important to ensure that the increase in FSD was not just due to underlying differences in facial expression.

Moderate or severe intellectual disability was present in 19 subjects with pathogenic SVs and 7 without pathogenic SVs. This is in keeping with the known association of SVs with intellectual disability (Helbig et al., 2014). Intellectual disability may affect FSD at the image capture stage. There may be less understanding of instructions given and more chance of images captured with the face in a different orientation or different expression. Nine out of 15 subjects with moderate to profound intellectual disability had their mouth open, for example.

Anti-epileptic drug history, according to number of drugs tried, was not significantly different between the two groups with and without pathogenic SVs after children were excluded. However, the same number of drugs had been used in those pathogenic SVs despite their mean age being eight years younger, suggesting that refractory epilepsy

may be more prevalent in people with pathogenic SVs. At the same time, the difference in age could mean that there is shorter duration of drug use in the younger group. However, there were no data on age of onset of epilepsy. The two drugs implicated in altering facial shape, sodium valproate and phenytoin, showed similar rates of use in both groups too. However, this is a limited analysis because there was no information available on duration or cumulative dosage of these drugs.

To some extent, the sensitivity analysis exploring weight may have corrected for the effect of sodium valproate and drugs such as pregabalin. On the other hand, it is difficult to justify excluding weight completely. It is well-known that some syndromes associated with epilepsy are also associated with obesity, such as Prader-Willi syndrome (Vendrame et al., 2010). Hence it is possible that some pathogenic SVs associated with epilepsy may also have an association with obesity. In this study, BMI was shown not to be a potential confounder. The role of BMI is explored further in Chapter 6 including a method of correcting FSD for BMI in this group of people with and without pathogenic SVs.

A major limitation of the study is the small sample size. In particular, only five patients with pathogenic structural variants were recruited for the validation set. Since I am investigating a group of conditions which are heterogeneous in terms of their effect on face shape, a larger sample of people with different SVs is necessary to show if FSD is as sensitive and specific as this study suggests. As noted earlier, the current sample was also not representative of the wider population with epilepsy. Thirty-five of those with pathogenic SVs had been selected for genetic testing based on clinical information which may have included a subjective impression of dysmorphism; the remaining 8 of the 43 with pathogenic SVs had been recruited and genotyped as part of a genome-wide association study.

Further research is needed to explore the usefulness of stereophotogrammetry and dense surface modelling, with larger patient samples, patients with other conditions, and other age groups and ethnicities. Other regions of the face, such as the ear, may also be informative in detecting dysmorphism. Finally, the role of potential confounding factors needs to be elucidated further. These include facial trauma, facial surgery, drug history and body mass index.

#### 4.6 CONCLUSION

This study shows that patients with pathogenic structural variants have a more atypical face. Structural variants are being discovered to be a more common cause of epilepsy and other clinical conditions than previously suspected, but it is not known which patients should be referred for clinical genetics testing, especially in adult medicine. I have shown that 3D stereophotogrammetry and dense surface modelling offers a promising avenue for detection of people who are likely to have clinically relevant structural variants. With further automation and refinement of stereophotogrammetry and dense surface modelling, it may help prioritise patients for genetic analysis for clinicians who are not experienced in dysmorphology or clinical genetics.

## CHAPTER 5: STRUCTURAL VARIANTS ANALYSIS

### 5.1 INTRODUCTION

Before the use of chromosome microarrays, SVs in people with a disease were considered to be rare. Karyotyping, under light microscopy, was the commonest clinical investigation for genome-wide detection of SVs, but is limited to a resolution of approximately 5 megabases (Sharp et al., 2006). An observational study of 24,901 amniocenteses performed for advanced maternal age found that 1.79% had chromosomal abnormalities on karyotyping (Caron et al., 1999). Population screening of newborns in Europe and Japan found prevalence rates of 0.85% and 0.63% respectively (Maeda et al., 1991; Nielsen and Wohlert, 1991). The majority of these were due to the following chromosomal disorders: trisomy 13, trisomy 18, trisomy 21, 45X, 47XXX, 47XYY, 47XXY and ring Y chromosome (Maeda et al., 1991).

SVs detected using karyotyping were always thought to be pathogenic, in that they were thought to be the cause of the disease (Shaffer and Lupski, 2000). Aneuploidies, such as Down syndrome, presented clinically with up to 100% penetrance, albeit with variable expressivity (Antonarakis et al., 2004), and even balanced translocations were associated with congenital malformations (Warburton, 1991). The identification of syndromes due to particular microdeletions (Shapira, 1998) and microduplications (Buckland, 2003; Lupski et al., 1991) using fluorescent in-situ hybridization (FISH), also suggested that submicroscopic SVs usually led to abnormal phenotypes.

It is now known that there is much greater copy number variation than previously recognised in the genomes of apparently healthy humans across different populations (Iafate et al., 2004; Redon et al., 2006; Sebat et al., 2004). Therefore, SVs cannot be assumed to be pathogenic and care is needed in interpreting results of chromosome microarray investigations.

Certain characteristics of SVs appear to predispose to pathogenicity (Girirajan et al., 2012). The overall number of SVs is correlated with pathogenicity. Deletions are more likely to be pathogenic than duplications. SVs that are de novo, instead of being inherited, and SVs that are larger or that contain more genes in the interval are more likely to be pathogenic.

In people with epilepsy, SVs are present in 5-30% depending on the study population, being more prevalent in the presence of features such as dysmorphism or intellectual disability (Helbig et al., 2014; Mullen et al., 2013; Striano et al., 2012). The SVs appear to be larger and contain more genes than in controls, especially for genes involved in ion channels and neurodevelopment (Lal et al., 2015; Striano et al., 2012). For example, SVs (especially deletions) that contain the genes for *SCN1A*, *KCNQ2* and *CHRNA4*, are associated with epilepsy (Kurahashi et al., 2009; Wang et al., 2008). Therefore, the functions of genes contained in an SV interval are also important in determining pathogenicity of an SV.

For many genes, the functions are not yet known or additional functions are recognised years after initial discovery. For example, the *LG11* gene was initially classified as a tumour suppressor gene associated with glial tumours and it was only later that it was found to be the cause of autosomal dominant partial epilepsy with auditory features (Kalachikov et al., 2002). For facial dysmorphism particularly, one alternative approach could be to look at gene expression in the relevant anatomical areas during the critical stages of facial development. Facial development is heavily dependent on embryonic forebrain genes (see Section 1.3.1). A transcriptome 'atlas' of embryonic and fetal gene expression in different parts of the human brain (Kang et al., 2011) enables a study of gene expression in people with atypical faces.

Databases are also used to help characterise the nature of a given SV. These can include internal databases for a laboratory. There are also external databases of normal

variants, such as Database of Genomic Variants, and of affected individuals, such as DECIPHER. It is thought more likely that a variant seen in supposedly 'normal' individuals will not be pathogenic, or that variants that appear in conjunction with certain disease phenotypes are more likely to be pathogenic. Sophisticated protocols for classifying SVs have been developed, incorporating SV characteristics and databases (Galizia et al., 2012) and more recently using genome-wide studies of cases and controls with SVs (Lal et al., 2015).

However, for many SVs, especially novel or rare SVs, it is not clear if they are indeed 'pathogenic' and such SVs may still be misclassified. In Chapter 4, it was assumed that SVs were pathogenic based on laboratory classification. As explained above, the basis for such classification is evolving and there is no universal agreement on the pathogenicity of all detected SVs. One study has suggested that 1 in 5 of purported causative SVs less than 500kb in size may not be pathogenic in some neurodevelopmental disorders, for example (Vermeesch et al., 2011). Others have suggested a two-hit hypothesis for developmental delay associated with SVs, stating that some SVs alone are not pathogenic but may be so in combination (Girirajan et al., 2010).

On the other hand, SVs may not be detected at all, or they may be misclassified as being normal variants when they are actually pathogenic. Methods of detecting SVs are varied and include older techniques, such as bacterial artificial chromosome clone-based assays and fluorescence in-situ hybridisation. Oligonucleotide arrays have also been developed, as well as algorithms to detect SVs using genotype data collected from single nucleotide polymorphism (SNP) arrays. Exome and whole genome sequencing have been used to detect SVs in people with epilepsy but have been thought to be less reliable for SV detection than oligonucleotide arrays (Leu et al., 2016).

Subjects in Chapter 4 had been genotyped using two methods, SNP arrays or aCGH. Also, aCGH was performed in two different clinical laboratories, which may have used different criteria. Therefore, pathogenicity of SVs had not been defined using uniform criteria.

In this study, I first look in more detail at the effect of SVs classified as 'pathogenic' in Chapter 4. This includes the use of a transcriptome 'atlas' to look for associations between gene expression and FSD. Then, I attempt to see if FSD can detect SVs that have not been classified as 'pathogenic'.

## 5.2 HYPOTHESES

I hypothesise that FSD is positively correlated with: a) the size of pathogenic SVs b) the presence of deletions instead of duplications and c) the number of genes expressed in the human embryonic forebrain. I also hypothesise that FSD will still be able to discriminate between people with pathogenic SVs and people without pathogenic SVs in a sensitivity analysis using method of SV detection (aCGH vs SNP array). Finally, I hypothesise that FSD will be increased in people with any detected SV greater than 500 kilobases, regardless of prior classification of pathogenicity.

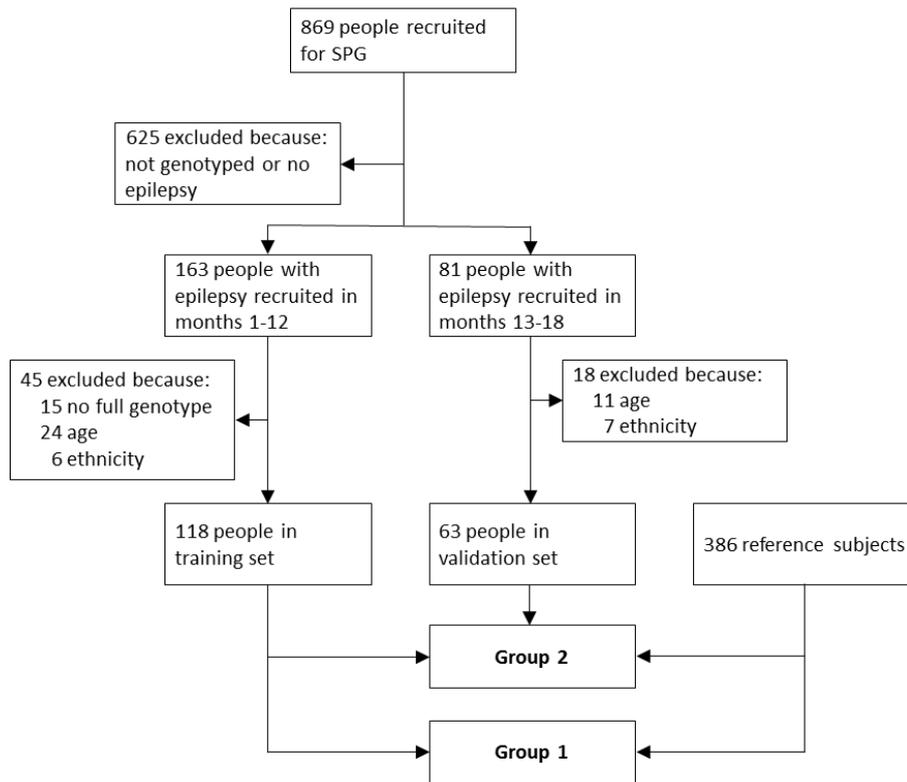
## 5.3 METHODS

Throughout this chapter, SVs are deemed to be pathogenic if classified as such by the laboratory reporting the SV. In my final hypothesis, I explore whether FSD can detect SVs that may not have been classified as pathogenic.

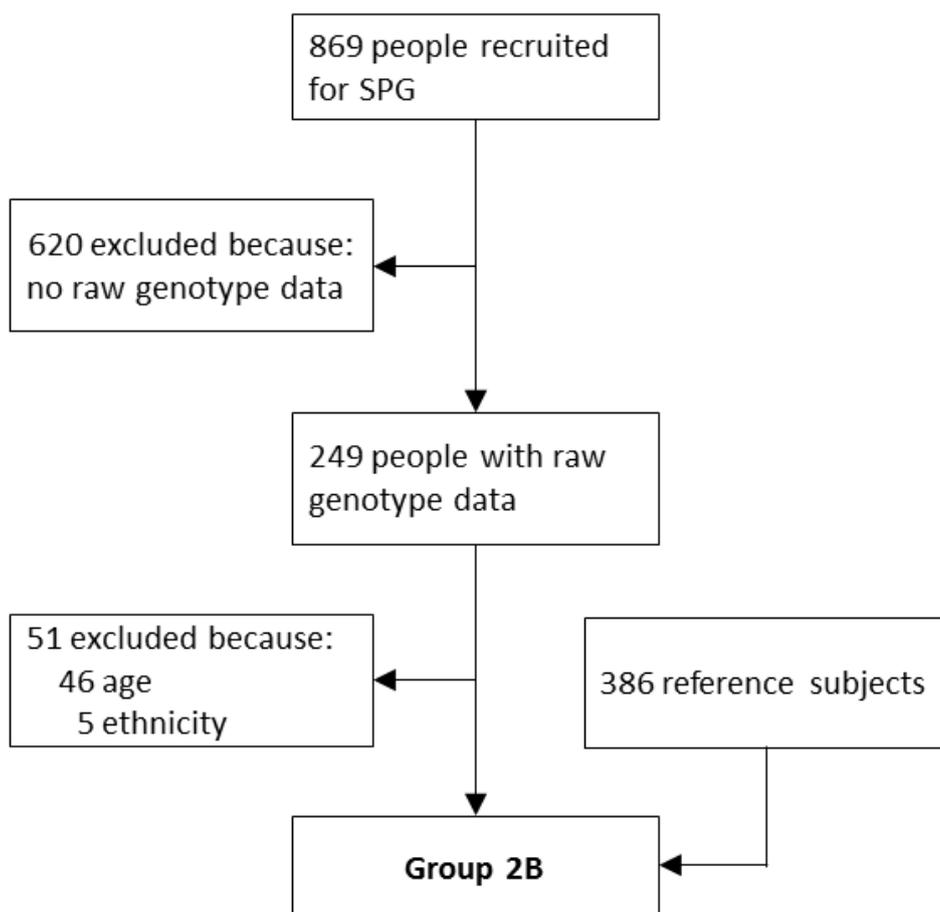
### 5.3.1 RECRUITMENT, IMAGE CAPTURE AND FACE SHAPE ANALYSIS

All participants in Group 2 were included in the studies of pathogenic SV size, gene content and forebrain gene expression. These subjects comprised children and adults who had been genotyped and had been diagnosed with epilepsy (Figure 46). Exclusion

criteria were those with monogenic epilepsies, those with known imbalances of a whole chromosome and those who could not be matched for FSD.



**Figure 46 Recruitment of subjects for Group 2, which comprised 181 genotyped children and adults with epilepsy. Of these, 43 had pathogenic SVs detected by aCGH or SNP array and 138 had no such SVs. Reference subjects were only used to calculate FSD. SPG = Stereophotogrammetry**



**Figure 47 Recruitment of subjects into group 2B.** This group comprised only people with epilepsy who had the complete genotype data available to allow assessment of all detected SVs, not just those that were pathogenic. 198 subjects were included after excluding those with no genotype data and those who could not be matched to reference subjects (due to age or ethnicity) for FSD calculation. SPG = stereophotogrammetry

For analysis of all SVs above a certain size, patients with epilepsy were recruited from a new group, Group 2B (Figure 47). In group 2B, only people with the complete raw genotype information available were included. This meant the exclusion of some subjects for whom the laboratory report provided information on all pathogenic SVs but no information was available for all other detected SVs. Group 2B included some people who had raw genotype information available but with no classification of pathogenicity; these people had undergone SNP array genotyping for a prior genome-wide association study. Other exclusion criteria comprised lack of age-matched

controls and lack of ethnicity-matched controls. Face images were collected as described previously in Chapter 3.

### 5.3.2 DATA COLLECTION

Genotype information was initially collected from different sources depending on the method used and is described in Chapter 5. More detailed SNP array data was collected from Dalia Kasperavičiūtė and Erin Heinzen, part of the EPIGEN research group which had initially performed the genotyping. The presence of SVs was assessed independently of a previous study using this data (Heinzen et al., 2010). Two people (Jan Novy and Vamsi Chinthapalli) used QuantiSNP (v2; Available from <https://sites.google.com/site/quantisnp/>), a Bayes hidden Markov model algorithm for detecting SVs from SNP array data, which has been previously shown to detect SVs from SNP array data (Colella et al., 2007). Default parameters were used to analyse data contained in Illumina GenomeStudio files. A threshold of at least 30 for the maximum log Bayes factor was used. During detection of SVs, copy number neutral loss of heterozygosity was ignored.

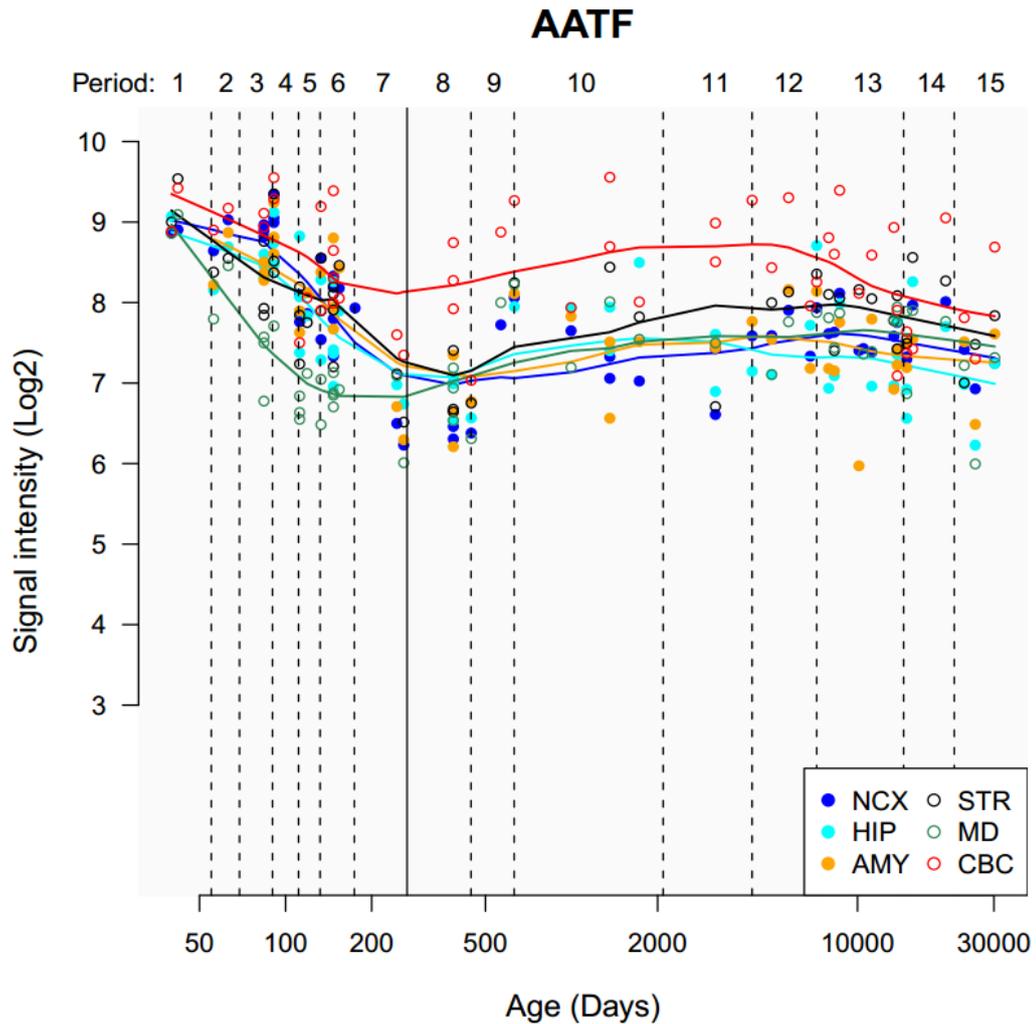
For aCGH, the full report was accessed from the relevant laboratories to obtain the details of all SVs. For UK patients, it was also possible to obtain the data regarding all other SVs, including those that were considered benign or of unknown significance. For those patients who had both aCGH and SNP array data available, only the aCGH data were used. Array CGH is more sensitive and specific than older SNP arrays (Greshock et al., 2007).

For the final hypothesis, tested in Section 5.4.4, all SVs greater than 500 kilobases were included in the analysis. This threshold was chosen based on previous reports, which suggest significant enrichment of SVs greater than this size in people with neurodevelopmental disorders (Vermeesch et al., 2011). In addition, the smallest

known recurrent microdeletion syndromes are approximately 500kb or larger in size (Girirajan et al., 2012; Miller et al., 2010).

### 5.3.3 BIOINFORMATICS ANALYSIS

The gene content of SVs was determined using the University of California, Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu>, hg version 18). The genome browser was accessed through the Genetic Diseases/Gene Discovery interface (GEDI; <http://gedi.ci.uchicago.edu/>) to obtain lists of genes contained within the affected interval. Genes adjacent to the SV interval were not included. Data on the breakpoints for all SVs were obtained from the original laboratories. Gene expression levels in the late human embryonic forebrain were acquired from the Human Brain Transcriptome database (<http://hbatlas.org/>). The Human Brain Transcriptome project collected transcriptome data in 27 regions of the brain from 57 post-mortem brain tissue samples in individuals with no known genetic disorders (Kang et al., 2011). They included developing brains with the earliest being at 5.7 weeks after conception, and there were two brain specimens used to establish gene expression under 56 days post-conception. Gene expression in tissue samples was found using the Affymetrix GeneChip Human Exon 1.0 ST Array and is displayed for each gene on a graph of log<sub>2</sub> signal intensity with time (Figure 48).



**Figure 48 Gene expression figure for a selected gene, AATF, from the Human Brain Transcriptome database. Signal intensity is plotted on a log base 2 scale. For this study, only genes expressed at a signal intensity >10 in the embryonic period were deemed to be expressed significantly (period 1, before 56 days post-conception). Therefore this gene was considered to be not expressed in the neocortex. NCX = neocortex, HIP = hippocampus, AMY = amygdala, STR = striatum, MD = mediodorsal nucleus of thalamus, CBC = cerebellar cortex.**

I noted only the earliest level of signal intensity in the neocortex, which was at or before 56 days' post-conception. This age was chosen as facial development in humans occurs mostly before seven weeks of age (Som and Naidich, 2013). Also, molecular signals from the brain are thought to exert the greatest influence on facial development in the early embryonic period, before 56 days' post-conception (Marcucio et al., 2015). Gene expression in the neocortex was categorised as present if signal intensity was

greater than 10 using a log2 scale or else as absent. This signal intensity threshold was used because genes known to be associated with cortical malformations, such as *DCX*, were expressed at a level greater than this (Kang et al., 2011).

#### 5.3.4 STATISTICAL ANALYSIS

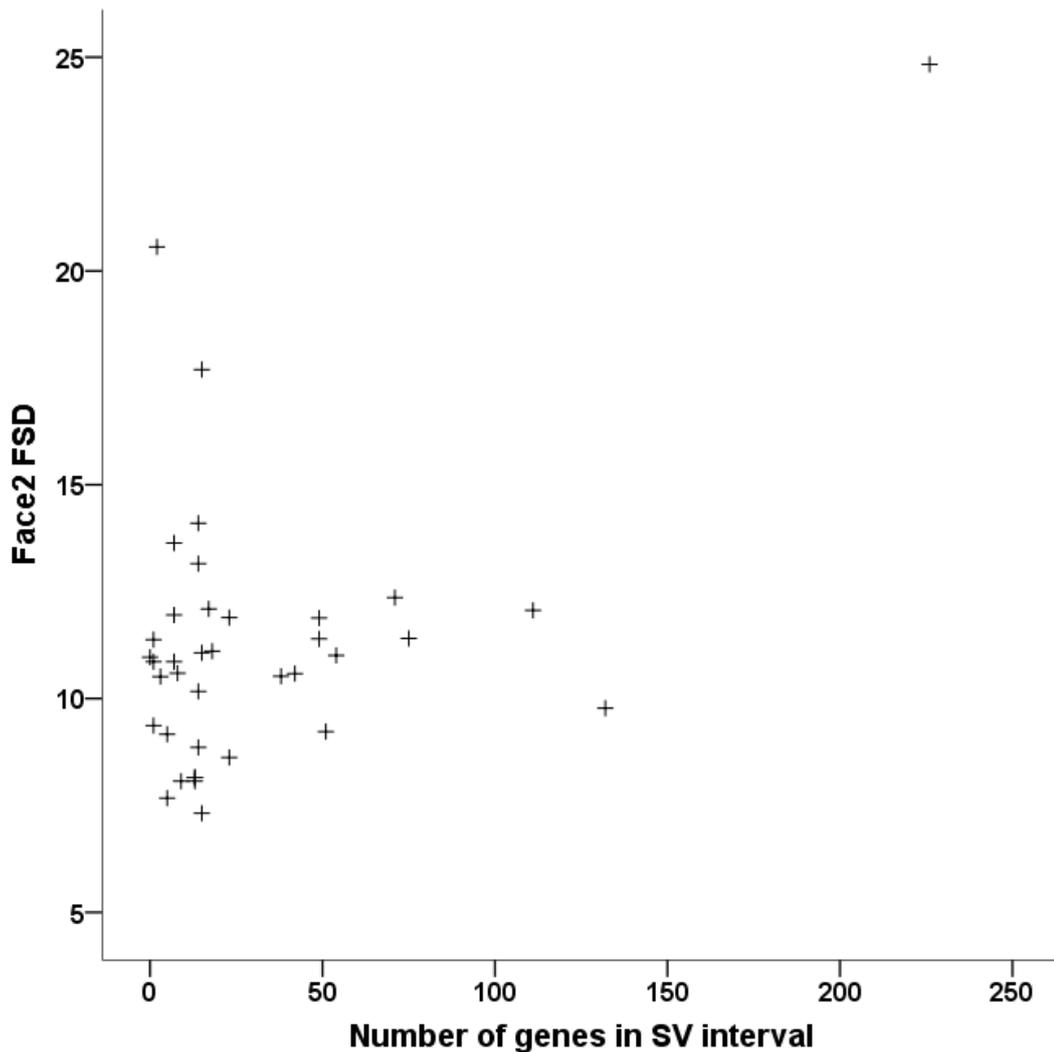
Spearman's rank correlation coefficient was calculated to assess the strength of correlation. For SV size and gene content, the variables were parametric and an independent samples t-test was used to compare differences. For FSD, non-parametric tests were used. For FSD in any DSM, the Mann-Whitney U test was used to compare differences in groups with and without SVs. The Kruskal-Wallis test was used to look for differences between more than two groups of people.

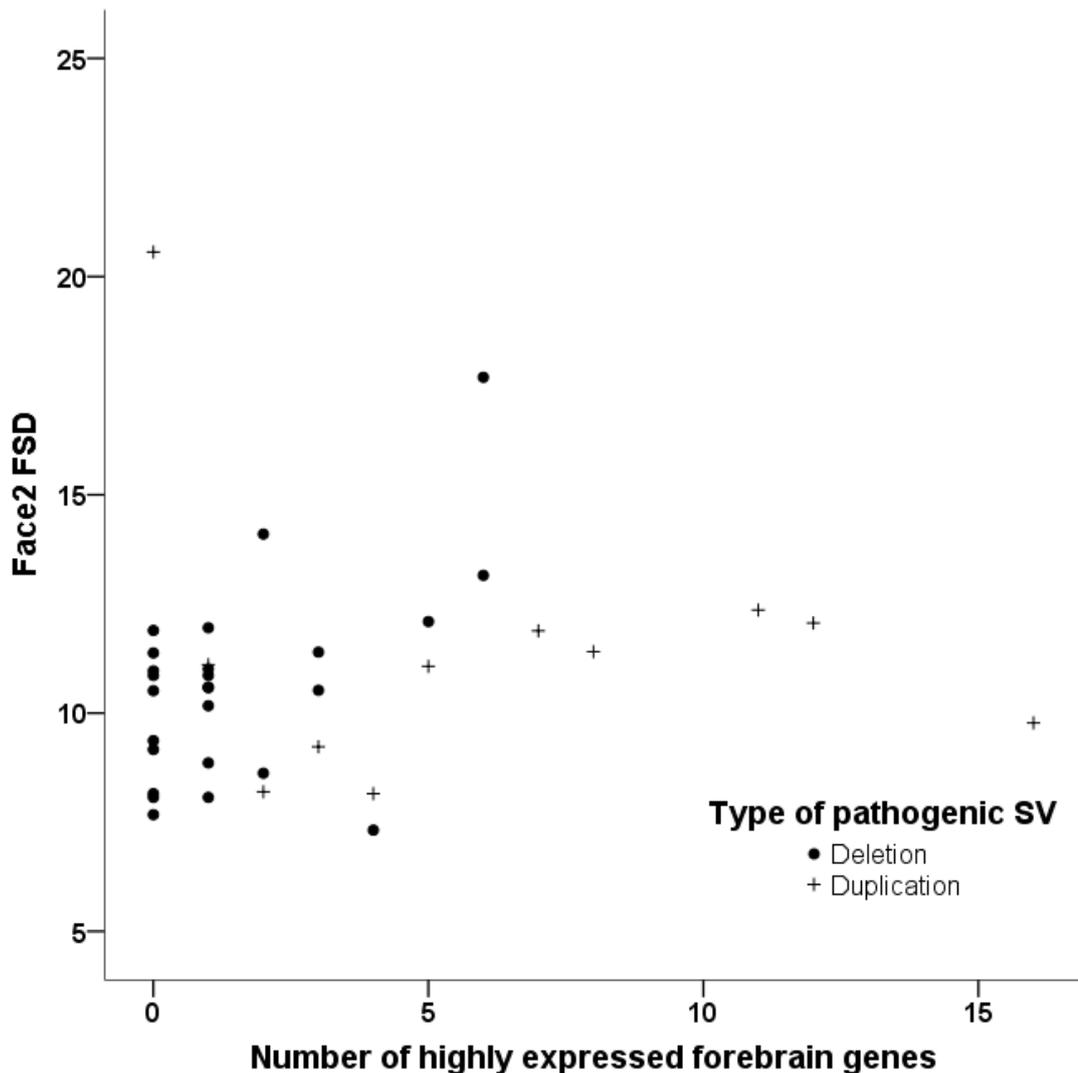
### 5.4 RESULTS

#### 5.4.1 PATHOGENIC SV SIZE AND GENE CONTENT

Three models incorporating both the training and validation cohort (Face2, Eyes2, Nose2) were used to explore the genomic and clinical data.

Exact breakpoints of the pathogenic SVs were known for 39 patients. Four others with pathogenic SVs had translocations or inversions detected by FISH or karyotyping, and so full data on pathogenic SV size were not available. They are not part of any subsequent analysis here. There was no correlation between whole face FSD and the size of the pathogenic SV in terms of total number of base pairs in the interval ( $\rho=0.20$ ;  $p=0.22$ ) or number of genes contained ( $\rho=0.19$ ;  $p=0.25$ ; Figure 49). Periorbital FSD showed positive correlation with the number of genes contained ( $\rho=0.38$ ;  $p=0.018$ ; significant after Bonferroni correction for 2 comparisons) but not the number of bases ( $\rho=0.25$ ;  $p=0.13$ ; not significant after Bonferroni correction for 2 comparisons).





**Figure 50** Scatter plot of the relationship between the total number of genes highly expressed in the embryonic human forebrain and within the breakpoints of a pathogenic SV for each patient, and FSD of the whole face. There was a significant positive correlation seen ( $\rho=0.34$ ;  $p=0.036$ ;  $n=39$ ). The correlation appeared to be stronger for patients with deletions than those with duplications.

In a secondary analysis, I looked at participants with deletions and duplications separately, and excluded two patients with both types of pathogenic SVs. Stronger positive correlation appeared to be present in patients with deletions ( $\rho=0.36$ ;  $p=0.07$ ;  $n=26$ ) than in those with duplications ( $\rho=0.15$ ;  $p=0.67$ ;  $n=11$ ). Mean pathogenic SV size and gene content were larger in patients with duplications than with deletions (11.5Mb vs 2.48Mb;  $p=0.009$ , 49 genes vs 13.5 genes;  $p=0.002$ ) but FSD was not significantly

different for any of the three DSMs between those with deletions and those with duplications.

#### 5.4.3 SENSITIVITY ANALYSIS PERFORMED BY METHOD OF SV DETECTION

Two genotyping methods for detecting pathogenic SVs were used: SNP arrays and aCGH. Thirty-one out of 43 patients (72%) with pathogenic SVs underwent aCGH, in contrast with only 15% of patients without pathogenic SVs. To exclude the possibility that method of SV detection is a confounder, I analysed only the people who underwent aCGH (n=56) and still found a significantly greater median FSD in those with pathogenic SVs using the whole face DSM (11.1 vs 9.26; p=0.005), or the periorbital DSM (13.7 vs 12.0; p=0.03), but not the perinasal DSM (9.11 vs 8.21; p=0.27).

A similar analysis could be performed for SNP array data. Overall, 125 people underwent SNP array only, of whom 8 had pathogenic SVs. Median FSD was not different in the whole face DSM (9.11 vs 8.63; p=0.38), the periorbital DSM (12.9 vs 11.4; p=0.38) or the perioral DSM (8.41 vs 7.82; p=0.22). The perinasal DSM showed significantly greater FSD in people with pathogenic SVs (8.81 vs 7.86; p=0.035).

Therefore, FSD, in subjects who underwent genotyping using SNP array, was greater in those with pathogenic SVs than in those without pathogenic SVs.

#### 5.4.4 FSD IN PATIENTS WITH ANY SV OVER 500KB

The final hypothesis was that FSD is increased in people with any SV, regardless of whether it has been classified as pathogenic or not. For this, a new dataset was used and analysis was performed in a new DSM. In total, 249 genotyped patients, with the full raw genotype, were recruited into group 2B (Figure 47). Of these, 46 were excluded due to being too old (>52years for men; >54years for women) for sufficient age-matched controls. Five people were excluded due to lack of ethnicity-matched controls and only white British adults were therefore included. Ninety-nine patients had at least

one SV >500kb, the cut-off value used, and in these subjects, the mean SV size was 1780kb (Table 31). There was no significant difference in FSD for any of the four DSMs between people with SVs and people without SVs.

	Patients with SV>500kb	Patients with no SV >500kb	Significance
Number	99	99	
Age	36.8	37.2	
Males	30	48	
aCGH	17	58	
SNP array	82	41	
Mean size (kb)	1780	N/A	
Face2B FSD	9.54	9.35	p=0.11
Eyes2B FSD	13.6	12.8	p=0.82
Nose2B FSD	8.9	8.84	p=0.73
Mouth2B FSD	6.91	7.21	p=0.26

**Table 31 Characteristics of patients with and without SVs >500kb and the median FSD for the whole face DSM (Face5), periorbital DSM (Eyes5), perinasal DSM (Nose5) and perioral DSM (Mouth5).**

Looking at the types of SVs, 67 subjects had duplications over 500kb and 18 subjects had deletions over 500kb. Another 14 subjects had at least one duplication and one deletion over 500kb.

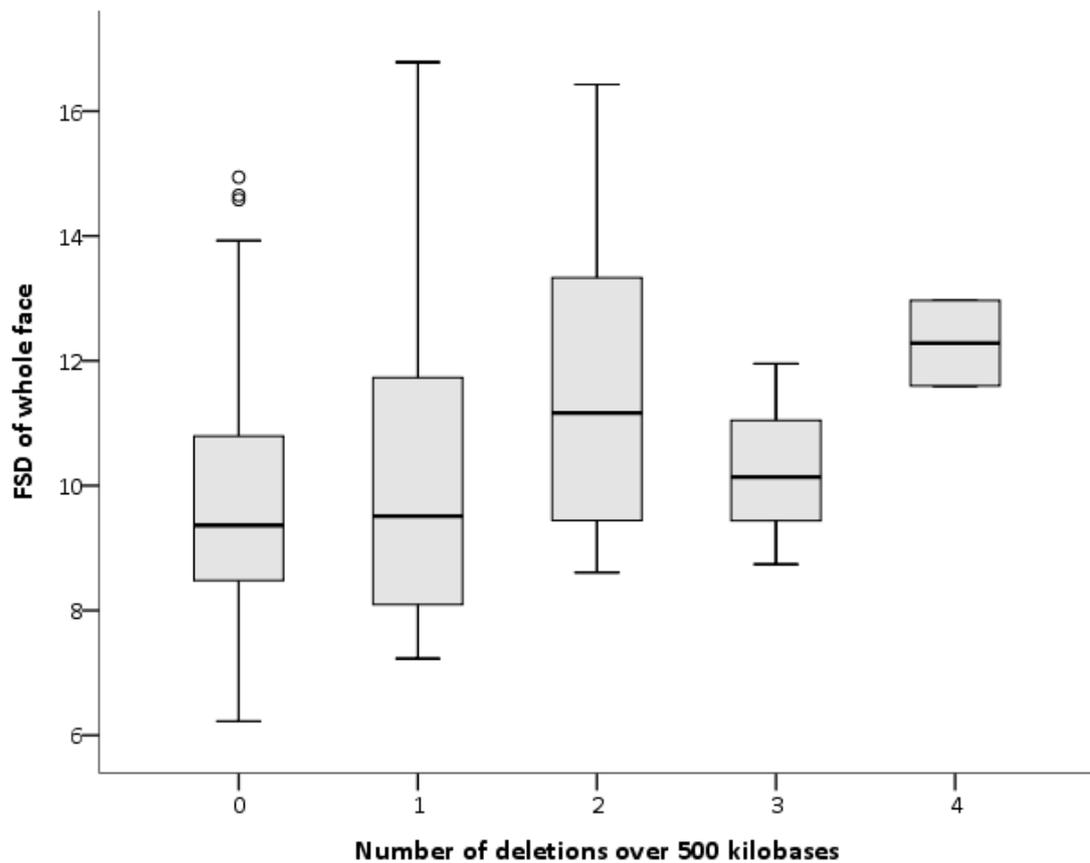
Though FSD was unable to detect a difference in the group of people with any SV over 500kb, it may be able to detect differences in people with deletions only. Deletions are more likely to be harmful than duplications (Girirajan et al., 2012) and phenotypic effects could include altered facial morphology. Therefore, FSD was compared between participants with any deletions over 500kb and genotyped participants without any known deletions over 500kb.

FSD appeared to be greater for the face images of those 32 people who had any deletions over 500kb compared to those people without known deletions. This was only significant for the perinasal region (Nose5: 9.48 vs 8.73; p=0.020; significant after Bonferroni correction for 2 comparisons) and not significant for the whole face (Face5:

10.5 vs 9.36;  $p=0.042$ ; not significant after Bonferroni correction for 2 comparisons), the periorbital region (Eyes5: 14.7 vs 13.0;  $p=0.029$ ; not significant after Bonferroni correction for 2 comparisons) or the perioral region (Mouth5: 7.76 vs 7.04;  $p=0.14$ ; not significant after Bonferroni correction for 2 comparisons). The results are shown in Figure 53. Hence, stereophotogrammetry and FSD may have a role in detecting deletions over 500kb. However, this is a small sample size and the results would need replication first.

In the previous chapter, I have shown that subjects with deletions and duplications labelled as pathogenic have an increased FSD. Some people in this group of 32 subjects with deletions also had deletions already determined to be pathogenic. So I looked to see if the increase in FSD was only because of those people who had deletions previously classified as pathogenic. In the 32 patients with any deletion over 500kb, 26 had undergone aCGH in a clinical laboratory: 11 of these were classified by the laboratory as having pathogenic deletions, 3 as deletions of unknown significance and 12 as having 'benign' deletions. Subjects with 'benign' deletions were compared to subjects with deletions classified as unknown or likely pathogenic significance and there was no significant difference in FSD for any DSM, although Face5 FSD was greater in the latter group (10.4 vs 11.4;  $p=0.46$ ). Therefore, the increase in FSD seen in people with any deletion over 500kb, might also be due to 'benign' deletions as well as pathogenic deletions in this small sample.

Fourteen of the 32 subjects had more than one deletion over 500kb. Rather than using just a binary classification for presence of deletions, I explored if the number of deletions over 500kb in an individual was associated with FSD. FSD appeared to show a positive trend in relation to the number of deletions for each individual (Figure 51). However, no significant difference was found in the whole face DSM (9.36 vs 9.51 vs 11.16 vs 10.14 vs 12.28;  $p=0.051$ ;  $n=32$ ; Kruskal-Wallis test).

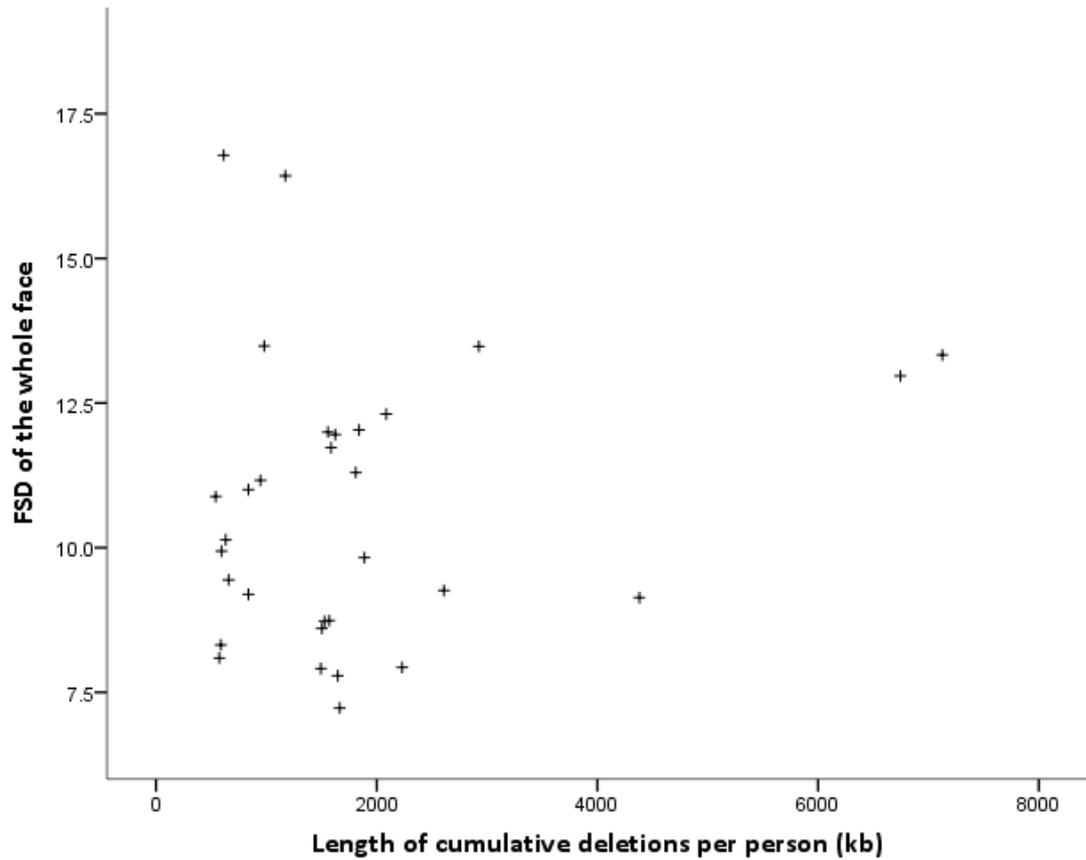


**Figure 51** The relationship between the number of deletions over 500 kilobases per person and the Face5 FSD. There was no significant difference in Face5 FSD between any of the groups (9.36 vs 9.51 vs 11.16 vs 10.14 vs 12.28;  $p=0.051$ ;  $n=198$ ; Kruskal-Wallis test)

I also looked to see if FSD was correlated with the cumulative length of all deletions over 500kb in subjects. There appeared to be weakly positive correlation but this was not significant (Figure 52;  $\rho =0.13$ ;  $p=0.48$ ;  $n=32$ ).

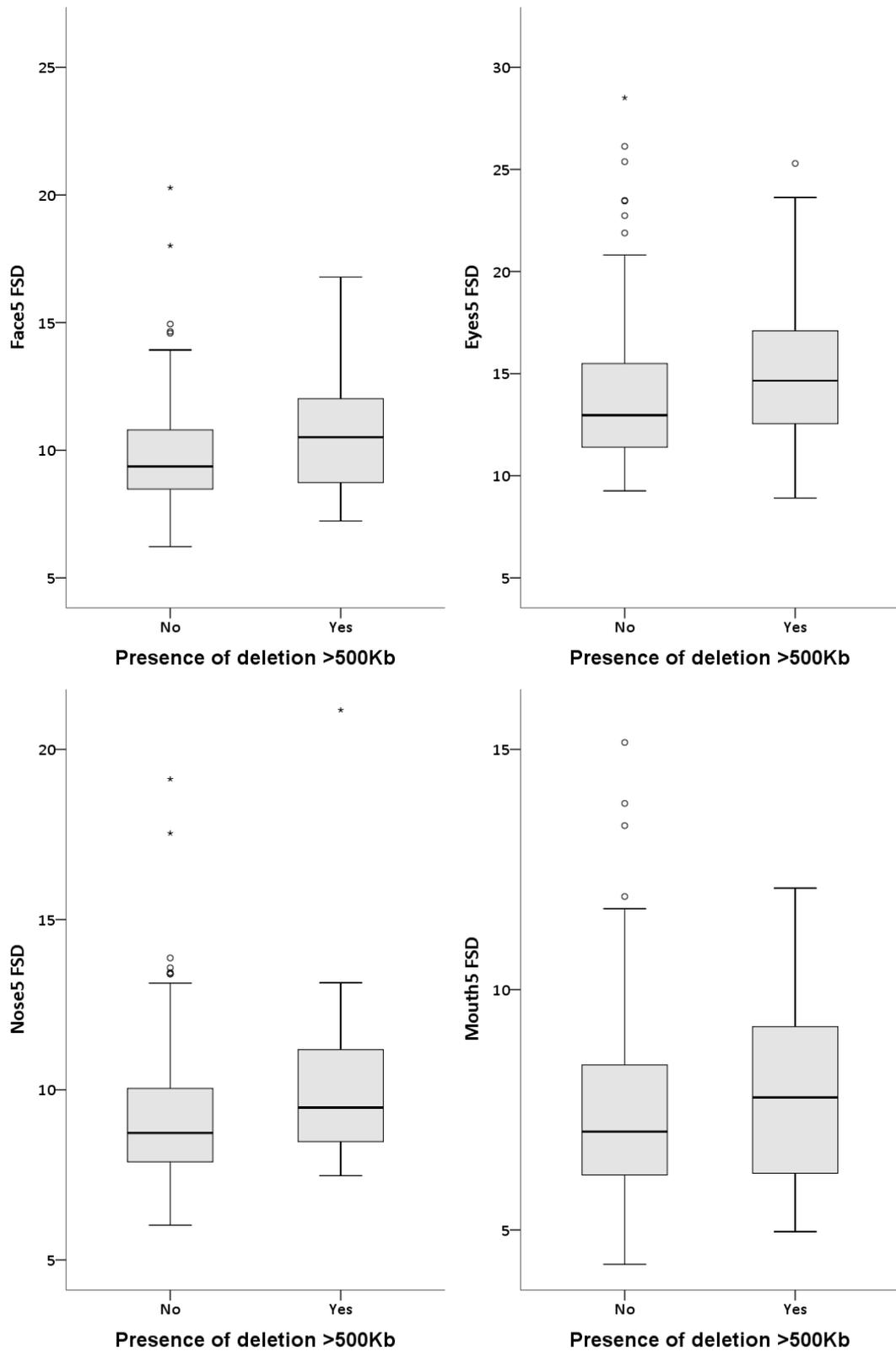
Overall, in all 32 people with deletions, there were 49 deletions. Twenty-one of these were ‘benign’ deletions that overlapped with at least one other ‘benign’ deletion. For example, there were overlapping deletions seen in chromosome 8p23.1 (chr8: 7,228,100-7,821,150, NCBI build 36.1, hg18), part of the beta-defensin gene cluster, which is known to show extensive copy number variation in humans (Machado and Ottolini, 2015). Excluding all overlapping ‘benign’ deletions, there were 28 deletions seen in 14 different chromosomes in 18 different subjects. Looking at just these 18

subjects, FSD of the whole face was still larger in the group with deletions but this was not significant (10.6 vs 9.36;  $p=0.11$ ;  $n= 184$ ).



**Figure 52** Scatter plot of FSD and the cumulative length of all deletions over 500kb found in an individual. There was no significant correlation between the two variables ( $\rho=0.13$ ;  $p=0.48$ ;  $n=32$ ).

In people with duplications only, there was no significant difference in FSD for any of the DSMs compared to people without duplications.



**Figure 53** Box plots showing the distribution of FSD in 4 DSMs in patients with deletions >500kb and patients without such deletions. The median FSD is greater for the perinasal region (Nose5: 9.48 vs 8.73;  $p=0.040$ ), but not for the whole face (Face5: 10.5 vs 9.36;  $p=0.084$ ), the periorbital region (Eyes5: 14.7 vs 13.0;  $p=0.058$ ) or the perioral region

**(Mouth5: 7.76 vs 7.04; p=0.28). Bonferroni correction had been applied for multiple comparisons.**

## 5.5 DISCUSSION

FSD was previously shown to differentiate between people with pathogenic SVs and people without pathogenic SVs. In this chapter, I further explored the relationship of FSD and SVs, both pathogenic SVs and non-pathogenic SVs. Initially, I looked at the size of pathogenic SVs and relationship to FSD. Only periorbital FSD was significantly correlated to the number of genes in an SV interval but this positive correlation was not strong ( $\rho=0.38$ ) and was not associated with the number of bases in an SV interval.

Whole face FSD was not associated with overall SV interval size in terms of number of genes or base pairs. SVs vary significantly in terms of gene content and the function of affected genes in an SV interval would be expected to be more relevant to the effect on face shape and FSD than size of SV alone. It may still be expected that overall SV burden is associated with both epilepsy and face shape, given that numerous genes and SVs have been associated with both phenotypes. A larger sample of people with SVs would be needed to look for such an effect.

To explore the function and importance of affected genes, a transcriptome 'atlas' of the human brain was used to find gene expression levels in the embryonic forebrain. For those genes that are contained within a pathogenic SV, there was weak positive correlation between the number of genes highly expressed in the forebrain at the embryonic stage and FSD. This analysis combined the use SV interval size with a 'functional' assessment of SV gene content, in which high gene expression in the embryonic forebrain was taken to mean a potential effect on facial embryology. Thus it is possible that genes in many of these SV intervals may be involved in regulating face and/or brain development. Some genes expressed in the human embryonic forebrain are already known to affect face development (Marcucio et al., 2015, 2011). One anthropometric study suggested that the *PAX3* gene is associated with the position of

the nasal root (Paternoster et al., 2012). Although there are methodological limitations explained in Chapter 1, the result is intriguing as *PAX3* is expressed in the embryonic murine midbrain and hindbrain, as well as being linked to a human condition, Waardenburg syndrome, with facial dysmorphism (Goulding et al., 1991). I found that *PAX3* was not highly expressed in the human forebrain when looking in the human transcriptome atlas. Genes thought to be potential causes of epilepsy syndromes, such as *CHRNA4* and *SCN1A*, were not necessarily expressed highly in the neocortex in the embryonic period, although genes implicated in epilepsy may exert their effects through later or more localised expression. The transcriptome atlas and relevance to genes involved in facial development is therefore not yet clearly understood. Reviewing all of the actual individual genes contained in an SV interval and their functions or clinical significance may be necessary in future studies.

Also, I showed that FSD is still able to differentiate people with pathogenic SVs from people without these SVs, when excluding SNP array cases altogether and only assessing aCGH cases. This is important because there is debate over the method of detection used for SVs, with SNP array potentially having a worse signal-to-noise ratio than aCGH (Alkan et al., 2011).

Apart from the platform used, a second difficulty in SV detection is in determining the pathogenicity of SVs. It is clear that this requires multiple steps and even then interpretation must be cautious (Fry et al., 2016; Galizia et al., 2012; Lal et al., 2015; Miller et al., 2010; Mullen et al., 2013; Olson et al., 2014; Striano et al., 2012). Methods used currently to determine pathogenicity of SVs in epilepsy can make more use of databases of known SVs, SVs containing known epilepsy genes and SVs in known epilepsy hotspots. Even so, some SVs are classified as likely pathogenic on the basis of size, inheritance and location.

In the final hypothesis of this chapter, previous interpretations of pathogenicity were ignored to see if FSD was useful in detecting people with SVs that may not have been deemed 'pathogenic'. All detected SVs greater than 500kb were considered to be relevant for analysis. This showed that for SVs overall, there was no change in FSD. However, for deletions over 500kb there was a significant increase in FSD in one out of four DSMs, which was the perinasal DSM. The other three DSMs suggested a trend to greater FSD in people with any deletion over 500kb, but this was not significant. Increased FSD was seen in some people with deletions over 500kb for deletions that had not been characterised as 'pathogenic' previously. If confirmed in further studies, this would suggest that FSD may be able to detect SVs that are thought to be benign, and thus these SVs may exert subtle effects on face shape that have been hitherto unrecognised. The implications of this may be broad, given that 5-10% of the general population are considered to have a SV greater than 500kb (Itsara et al., 2009).

Inspecting these 'benign' deletions, some were in areas of known copy number variation, such as those in the beta-defensin gene cluster and these were excluded. After exclusion, the number of people with deletions was only 18 and there was no significant increase in FSD. However, even these areas of known copy number variation have been found to be linked to clinical phenotypes, with, for example, copy number variation in the beta-defensin gene cluster being linked to Crohn's disease, psoriasis and HIV susceptibility (Machado and Ottolini, 2015). Hence, the implications of greater FSD in people with any deletion over 500kb are not clear. There may be confounding from deletions with no clinical significance or there may be as yet unknown effects on facial development and morphology from such deletions.

A limitation of the study is the small sample size, especially in sensitivity analyses and in comparison to other studies of SVs in people with epilepsy (Helbig et al., 2014; Mefford et al., 2010; Mullen et al., 2013; Striano et al., 2012). Also, the techniques used

here including the imputation of SVs from SNP array data and determination of significant SVs using size only are useful for an exploratory analysis only.

The bioinformatics analysis of the genes contained in an SV interval is also limited both in the genes considered and the extent of gene function. Genes adjacent to the SV interval were ignored. However, it could be argued that such genes may be affected by changes in copy number in neighbouring DNA through a number of mechanisms (de Smith et al., 2008). A gene that overlaps the SV breakpoint could also have reduced expression or may form a new fusion gene. Certain SVs are known to affect the expression of flanking genes (both upstream and downstream) through disruption of regulatory elements necessary for those genes. There may also be position effects in which a regulatory or promoter element may be closer to or further from its target gene due to an SV, which could then lead to increased or decreased gene expression, respectively. Structurally, an SV may cause some neighbouring genes to be located closer to a heterochromatin region of DNA, which is more tightly packed and so potentially more difficult for gene expression to occur within it.

On the other hand, it is not always the case that genes within an SV interval are affected uniformly (de Smith et al., 2008). Some SVs may occur in chromosomal loci that exhibit imprinting: In this case, the effect of the SV in terms of gene expression would depend upon whether it occurs in the paternal or maternal copy. Some genes, such as those for  $\alpha$ -defensins, show no effect of copy number variation on mRNA levels.

In terms of the genes studied for the bioinformatics analysis, I avoided any assumptions about the gene function and used a transcriptome atlas to look at the expression levels in the developing human brain. However, different time points or sites of expression could have been chosen. Another approach for studying the genes is pathway analysis using raw genotype data. This has the advantage of potentially detecting changes due to

an SV in a known molecular pathway and thus inferring a possible mechanism by which facial dysmorphism occurs.

Another limitation was that FSD-based analysis was limited by the lack of age-matched controls above the ages of 52-54 years. This meant that about 40% of the genotyped patient cohort could not be analysed. Incorporating more controls would help increase sample size in future studies using FSD and DSMs.

## 5.6 CONCLUSION

FSD is significantly increased in people with SVs, particularly deletions, after correcting for the potentially confounding effects of method of SV detection or laboratory-determined pathogenicity. FSD is not associated with overall size of SVs though. FSD also appears to be associated with genes expressed in the human forebrain and appears to be able to detect deletions of no apparent clinical significance. These initial findings need to be addressed in larger studies.

## CHAPTER 6: FACE SHAPE AND THE EFFECT OF BODY MASS INDEX AND AGEING

### 6.1 INTRODUCTION

Individual human body shape changes with body weight. The widely-used Body Mass Index (BMI) adjusts weight (kg) to the square of height (m<sup>2</sup>), and was considered a good predictor of skinfold thickness and body 'build' (Keys et al., 1972). BMI shows a significant correlation with the circumference of the head, chest, waist, hip, arm, thigh and knee (Wells et al., 2007), and with the profile area of the chest, waist, and hip (Lin et al., 2002). BMI correlates strongly with total fat mass, abdominal subcutaneous fat mass, and abdominal visceral fat mass measured by MRI (Janssen et al., 2002).

Clinically, increased BMI has been used as a measure of obesity, and predicts increased risk of cardiovascular disease, cerebrovascular disease, cancer and mortality across different ethnicities (Whitlock et al., 2009; Zheng et al., 2011).

The effect of BMI and body weight on craniofacial shape remains unclear. Research has focussed on the predisposition to obstructive sleep apnea (OSA). BMI is associated with increased neck circumference, thyromental angle, and thyromental distance in patients with OSA (Lam et al., 2005), but it is suggested that craniofacial characteristics contribute to OSA largely independently of BMI (Lee et al., 2010). Only one group has looked at facial soft tissue changes in OSA, and reported that Malay patients with OSA had increased BMI and increased buccal and submandibular fat deposition (Banabilh et al., 2009).

Knowledge of the facial changes that occur with a changing BMI could help better refine the technique of DSM and FSD. This is important because in Chapter 4, I found that there is a weak positive correlation of FSD with BMI. A number of genetic syndromes, such as Bardet-Biedl syndrome, Prader-Willi syndrome and Smith-Magenis syndrome,

are also associated with obesity (Cassidy et al., 2012; Elsea and Girirajan, 2008; Mykytyn et al., 2002). Notably all three of these disorders have been analysed using dense surface modelling (de Souza et al., 2013; Hammond et al., 2005; Tobin et al., 2008). Epilepsy is also associated with some obesity syndromes, including Prader-Willi syndrome and Smith-Magenis syndrome, and obesity may therefore be linked to some genetic causes of epilepsy.

In addition, it may also predict those susceptible to conditions such as OSA or nocturnal asthma (Calhoun, 2003). Facial fat distribution may also be a useful marker of cardiovascular risk and mortality, similarly to anthropometric abdominal fat measurements (Ashwell et al., 2012). This could negate the need to use whole-body imaging methods, as have been proposed (Wells et al., 2007). Therefore, the first part of this chapter explores the effect of BMI on FSD in a larger sample of people with epilepsy than in Chapter 4. I also investigate whether the effect of BMI on FSD can be controlled for.

A second part of this chapter was to assess the effect of ageing on FSD in particular and face shape in general. Much of the research into the effects of ageing on the face are from the field of orthopaedics. Since the 1920s, it has been known that there is a gradual increase in the size of the craniofacial skeleton throughout life, with, for example, a 5% increase in the size of the skull between adolescence and old age (Israel, 1973). Longitudinal cephalometric studies over the last 40 years have confirmed skeletal growth in adulthood and have also shown that there is an increase in facial height and chin protrusion with age (Akgül and Toygar, 2002; Forsberg, 1979; Pecora et al., 2008; West and McNamara, 1999). Lateral cephalograms also delineate the soft tissue profile of the face and these studies noted face surface changes. For example, Pecora and colleagues have drawn the face in profile at three different time points over 40 years for male and female subjects (Pecora et al., 2008). They found that there was

an increase in lip length but a decrease in upper lip thickness. Soft tissue thickness increased in the chin and the nose grew forwards and downwards throughout the time period up to at least a mean age of 57-58 years.

Farkas and colleagues used direct anthropometry in the largest study of age-related changes to the face in 600 adults, although this was a cross-sectional study (Farkas et al., 2004). They showed that facial width (zygoma – zygoma) was relatively unchanged between different age groups in adults. By contrast, jaw widths and facial height increased by approximately 1-5mm into middle adulthood, defined as 50-60 years, with a subsequent decrease by the age of 80 years, possibly due to muscle atrophy. Direct anthropometry using handheld digitizer probes has been used to measure specific nose or lip distances in two other cross-sectional studies which confirm earlier cephalometric findings (Sforza et al., 2011, 2010). These studies again highlight a limitation of anthropometry because they were unable to comment on shape changes in regions without landmarks and detected shape changes are described in terms of an increase or decrease in inter-landmark distances. More recently, stereophotogrammetry has been used in plastic surgery to assess facial skin sagging due to gravity (See et al., 2008). A recent review summarises what is known about facial ageing in different regions for each layer from skin to bone (Gerth, 2015).

With dense surface models (DSMs), age has been described as an ‘unavoidable’ confounder in studies and the effect of age on facial features in certain syndromes has been explored (Hammond et al., 2005, 2004). In all DSMs using a wide age range of subjects, the first PC seems to represent overall facial size in terms of both height and width. However, these findings from DSMs have mostly been from studies including children and adults, and have all been cross-sectional. There has been no longitudinal study using dense surface modelling to assess the effect of facial changes with age on face shape analysis. This is important because if FSD is to be used as a clinically useful

marker, it should not be unduly affected by age. Also, facial dysmorphism is known to vary in some syndromes with age and can become more obvious or less obvious with age. For example, 22q11 deletion syndrome is found to become less obvious with age, in a study using dense surface modelling (Hammond et al., 2005). Age-matching should account for variation due to age, but this has not been proven in any study.

Here I use stereophotogrammetry and dense surface modelling to explore changes in face shape with weight, height, BMI and ageing in a group of people attending an epilepsy unit.

## 6.2 HYPOTHESES

Firstly, I hypothesise that stereophotogrammetry and dense surface models can identify variation in face shape due to differences in body mass index. Secondly, I also hypothesise that this is predictable and thus the effect of body mass index on FSD may be reduced. Thirdly, during follow up of patients undergoing repeat stereophotogrammetry, I hypothesise that in a time interval of up to two years, there is no significant difference in face shape analysis, including FSD.

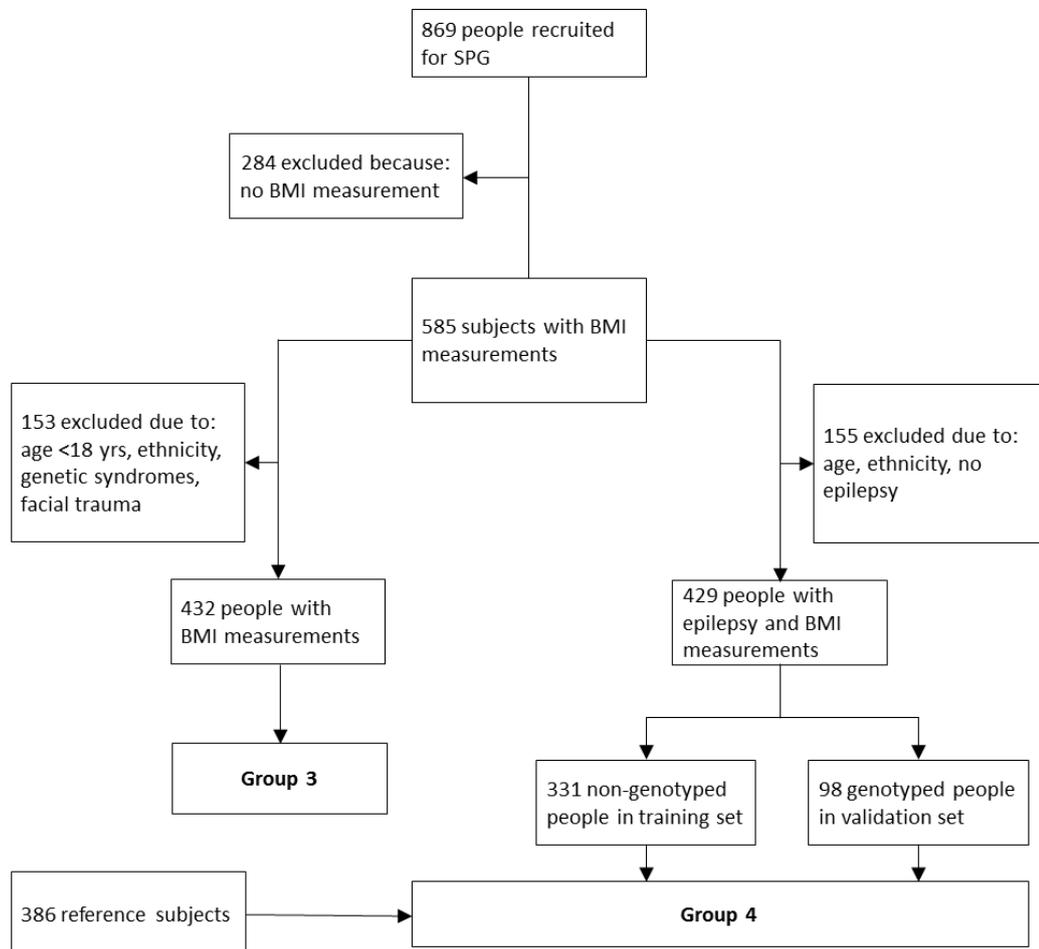
## 6.3 METHODS

### 6.3.1 RECRUITMENT, IMAGE CAPTURE AND FACE SHAPE ANALYSIS

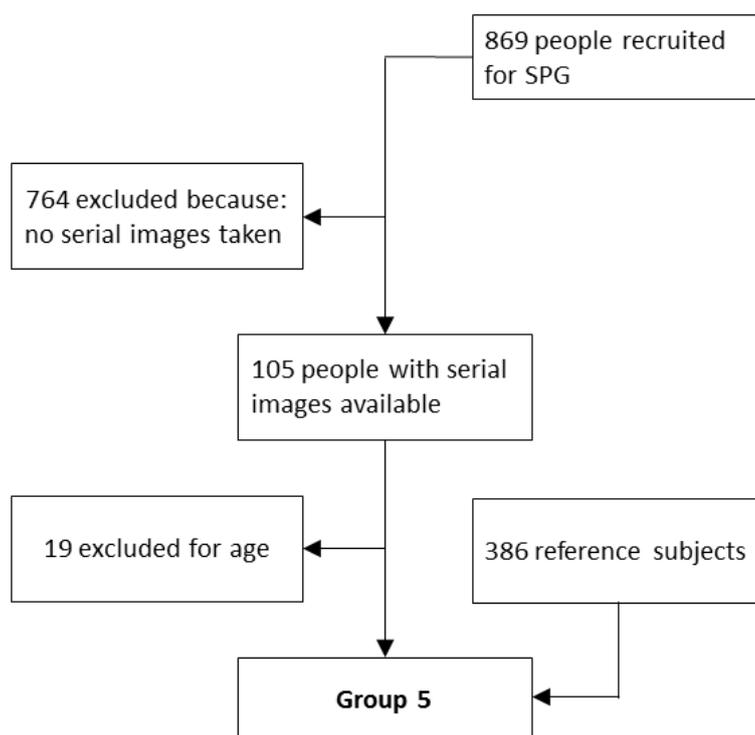
Patients with epilepsy and relatives at two UK centres were recruited as described for groups 3, 4 and 5 in Chapter 2. In short, group 3 comprised adults with height and weight measurement, excluding potential confounders. Such confounders included known genetic syndromes (including those with SVs) and facial trauma. This group was chosen to establish and describe the relationship of BMI to face shape in DSMs. The exclusion criteria were chosen to minimise confounding by other factors.

Group 4 was used to test the second hypothesis that the effect of weight could be minimised for FSD. FSD itself could not be tested in group 3 because that group

contained no reference subjects, included people without epilepsy and also had people who would not have sufficient matched reference subjects to enable FSD calculation. Group 4 comprised only adults with epilepsy who had height and weight measurements. Additionally, people were only included if they could be matched for age, sex and ethnicity to sufficient reference subjects to allow FSD to be calculated. The recruitment of groups 3 and 4 is shown in Figure 54. Group 5 comprised people with epilepsy, who underwent repeat imaging of the face after recruitment during subsequent visits to the UK site for clinic appointments (Figure 55).



**Figure 54** This is a duplicate of Figure 8. It summarises the characteristics of those recruited for Group 3 and Group 4. Group 3 was for an initial exploratory study of whether face shape was affected by BMI after excluding 153 potential confounders. Group 4 used less strict criteria but only recruited people with epilepsy. Group 4 was used to see if SVs in genotyped patients could be detected. SPG = Stereophotogrammetry



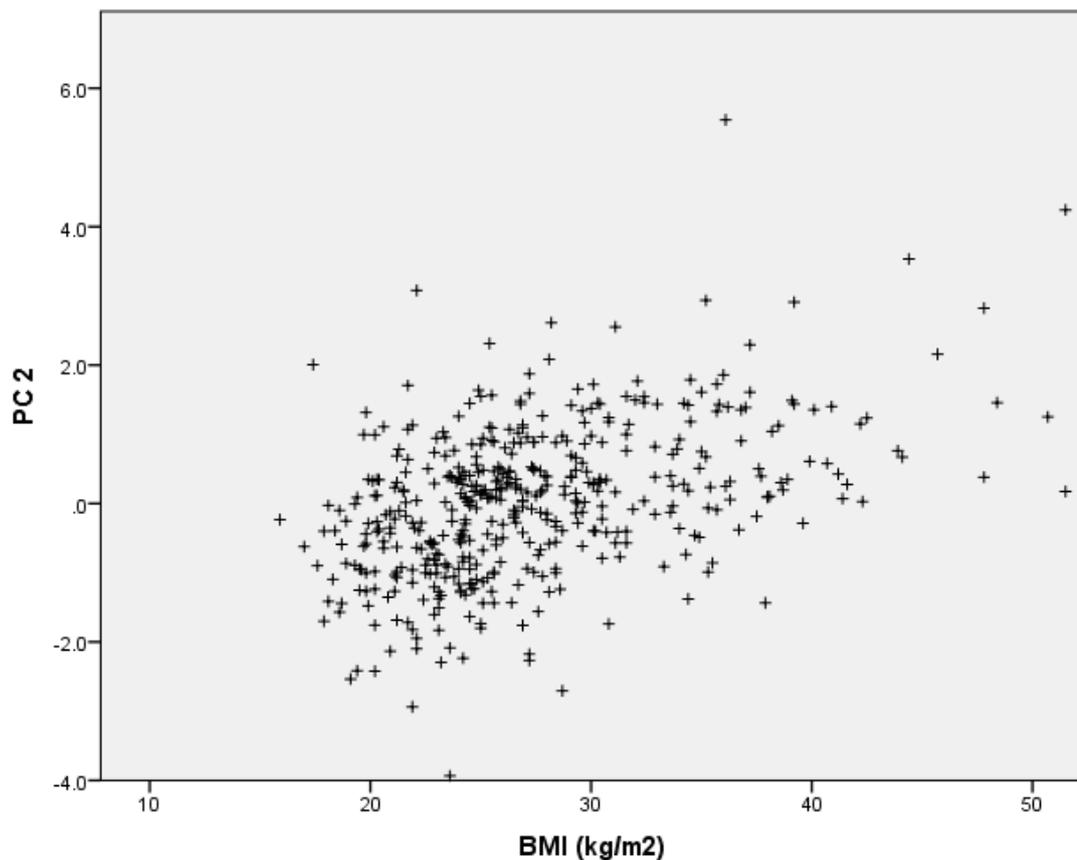
**Figure 55 Recruitment of subjects for Group 5, a study of face shape change with age. All people with epilepsy, with serial images taken were included in this study.**

Height and weight were recorded as in Chapter 2. BMI was then calculated and when BMI is referred to later in this chapter, the units are in  $\text{kgm}^{-2}$  unless otherwise stated. Group 3 was used to explore the relationship between principal components (PCs) and BMI. Group 4 was used to test whether the effect of BMI could be minimised. Group 5 was used in four DSMs to see if there are changes to FSD and face shape over time.

### 6.3.2 STATISTICS

In testing the first hypothesis, group 3 DSMs were used. Spearman's rank correlation coefficient was calculated for associations. Since some data were not normally distributed, I used the Kruskal-Wallis non-parametric test to compare different BMI groups. Then standard multiple regression was applied to assess the strength of any relationships, with a minimum ratio of 5 cases per independent variable. DSMs use multiple PCs to describe facial shape and there were 81 PCs for Face3 (whole face), 172 PCs for Eyes3, 70 PCs and 79 PCs for Mouth3, used to describe 99% of the total

variation in shape. To maintain 5:1 ratio, as per standard multiple regression assumptions, some PCs were not assessed. PCs were selected consecutively from the first PC in all cases until the 5:1 ratio was reached. This is because successive PCs explain less variation in a face shape than the ones preceding them. For those PCs associated with BMI, tests of ANOVA deviation from linearity were performed to check that PCs varied linearly with BMI (Figure 56).



**Figure 56** An example of the relationship between one principal component (PC 2) and body mass index (BMI) in adult subjects with height and weight measurements, who formed group 3. Testing for deviation from linearity showed that the relationship was significantly likely to be linear.

For the second hypothesis, which was to assess if the effect of weight and BMI on FSD was predictable, standard multiple regression in a DSM of group 4 confirmed that PCs 1 to 4 were significantly associated with BMI in this DSM too as expected. It was important to reassess this as PCs are not the same in different DSMs. Then group 4 was divided into two subgroups, for training and validation. Validation was to see if

adjustment for BMI affected the difference in FSD between people with SVs and people without SVs. The training set comprised people who had not been genotyped. The validation set was of genotyped people with known BMI, and either had SVs or no SVs. Therefore, using formulae developed from the training set, I aimed to adjust face shapes in the validation set to a standard BMI and see how this affects FSD. Specifically, the purpose was to see if FSD was still significantly different between people with SVs and people without SVs, once FSD was adjusted for BMI. All of the genotyped people in the validation set had been part of Group 2, the group of all genotyped people with or without height and weight measurements, which was studied in Chapters 4 and 5.

Simple linear regression with BMI as an independent predictor was used to quantify the change in PCs 1 to 4 for each unit of BMI in all subjects not part of group 2, for each sex separately. The remaining PCs were not altered. These 4 PCs were then adjusted for weight in the calculation of an adjusted FSD value according to the following formula:

$$\begin{aligned} \text{FSD}(x_{adj})^2 = & \Sigma(x_{PC1} - m_{PC1} - ((w_x - 20) \times b_{PC1}))^2 \\ & + (x_{PC2} - m_{PC2} - ((w_x - 20) \times b_{PC2}))^2 \\ & + (x_{PC3} - m_{PC3} - ((w_x - 20) \times b_{PC3}))^2 \\ & + (x_{PC4} - m_{PC4} - ((w_x - 20) \times b_{PC4}))^2 + (x_{PC5} - m_{PC5})^2 \\ & + (x_{PC6} - m_{PC6})^2 + \dots \end{aligned}$$

where  $x$  is an arbitrary subject's face which is compared to  $m$ , its matched mean from the control set,  $x_{adj}$  is the face adjusted for a BMI of 20.0kg/m<sup>2</sup>,  $w_x$  is the BMI of subject  $x$ , and  $b$  is the gradient in the linear equation of the relationship of BMI with a particular PC (for PC1 to PC4).

This formula is similar to the original formula used to calculate FSD except that the terms for the first four PCs include:  $((w - 20) \times b)$ . This term is an adjustment made for

weight using the previously derived linear regression equations for the first four PCs.

The term is still valid if the subject's BMI ( $w$ ) is less than 20 kgm<sup>-2</sup>.

For example, for the first male subject, his BMI was 39.2kg/m<sup>2</sup> and the coefficient for PC1 was 0.053 for males. Therefore, PC1 was reduced by  $0.053 \times (39.2 - 20.0)$  to adjust it for a BMI of 20.0. PC1 for this subject was then subtracted by his matched mean subject's PC1. This was repeated for PC2, PC3 and PC4. FSD was then recalculated with the new BMI-adjusted values for PC1-PC4 for all subjects in the group. Therefore, all of the PCs associated with BMI changes are now adjusted for each subject's BMI, and all other PCs are not changed.

In the validation set of genotyped people of known BMI, the Mann-Whitney U test was used to test for a significant difference in BMI-adjusted FSD between those with pathogenic SVs and those without pathogenic SVs.

For the longitudinal study of facial ageing, differences in FSD and BMI between the set of first and second images of subjects were compared using the related samples Wilcoxon signed rank test. Spearman's rank correlation coefficient was also measured for the difference in FSD with other variables. Bonferroni correction was applied for multiple comparisons.

Analysis was conducted using SPSS (versions 19-23, SPSS Inc; Chicago, IL, USA).

## 6.4 RESULTS

### 6.4.1 SUBJECT POPULATION

For group 3, the group of people of known BMI, I recruited 432 subjects in total, of whom 373 were patients and the remainder were unaffected relatives. Table 32 summarises their characteristics grouped by sex and BMI. The mean BMI was not significantly different between men (27.4; n=196; 45%) and women (27.3), but men had a greater mean age (44.4 vs 40.1 years; n=432; p=0.001). I divided people by BMI

into 4 groups: underweight (BMI<18.5), of normal weight (BMI 18.5 – 24.9), overweight (BMI 25 – 29.9) or obese (BMI >29.9). For each sex, height and age did not differ significantly between the BMI groups.

Relatives were significantly older than patients (55.3 vs 39.9 years; p<0.001), and 41 of the 59 were parents of patients with epilepsy. Relatives did not have a significantly different height, weight or BMI to patients.

BMI categories (kgm <sup>-2</sup> )		<20	20 – 24.9	25 – 29.9	≥30	All
<b>Male subjects</b>						
Total	(n)	10	57	77	52	196
Patients	(n)	10	52	67	43	172
Relatives	(n)	0	5	10	9	24
Age	(yrs)	42.3	41.4	45.9	45.8	44.4 (14.1)
Height	(m)	1.77	1.79	1.76	1.75	1.77 (0.08)
Weight	(kg)	59.7	74.5	84.1	104.4	85.4 (16.1)
PC 1		0.268	0.744	0.906	1.225	0.911 (0.457)
PC 2		-0.792	-0.374	0.195	0.985	0.189 (1.144)
PC 3		-0.915	-0.316	0.118	0.522	0.046 (1.017)
<b>Female subjects</b>						
Total	(n)	20	90	62	64	236
Patients	(n)	19	75	53	54	201
Relatives	(n)	1	15	9	10	35
Age	(yrs)	38.3	38.1	42.3	41.3	40.1 (13.6)
Height	(m)	1.62	1.65	1.63	1.62	1.63 (0.07)
Weight	(kg)	50.1	61.7	71.7	96.4	72.8 (18.6)
PC 1		-0.393	-0.199	-0.048	0.255	-0.053 (0.453)
PC 2		-0.672	-0.429	-0.014	0.417	-0.111 (0.973)
PC 3		-0.414	0.049	0.143	0.504	0.158 (1.051)

**Table 32 Characteristics of all 432 participants, grouped by sex and BMI category. The majority of participants were patients (n=373), not relatives (n=59). All other values shown are the means (standard deviation in parentheses). Height and age were not significantly different in people of different BMI. Using multiple regression, PCs 1 to 3 in the whole face DSM were significantly associated with BMI and weight in each sex, and they are seen to increase significantly between BMI groups.**

#### 6.4.2 FACIAL SHAPE AND ASSOCIATION WITH BMI AND BODY WEIGHT

For the whole face DSM, 81 PCs were used, which accounted for 99% of the variation in facial shape. Men and women were assessed separately. In men (n=196), I used the first 39 PCs, which explain 98.1% of facial shape, in a standard multiple regression analysis.

Using a maximum of 39 PCs maintains the minimum 5:1 ratio of cases to independent variables. The coefficient of determination ( $R^2$ ) was 0.709 for BMI and 0.713 for body weight (Table 33). In women (n=236; 5:1 ratio), regression analysis using the first 47 PCs (98.3% of facial shape) resulted in a coefficient of determination of 0.701 for BMI and 0.713 for body weight. All coefficients were highly significant ( $P<0.0001$ ). When looking at individual PCs in the regression model, only PCs 1 to 3 were significantly associated with BMI ( $P\leq 0.001$ ) and weight ( $P<0.05$ ) in both sexes.

The three other DSMs looked at different face regions, and none of these were as strongly associated with BMI or weight as the whole face (Table 33). The periorbital DSM gave the highest coefficients of determination with weight ( $R^2=0.695$  in men; 0.666 in women) and BMI ( $R^2=0.643$  in men; 0.636 in women). The relationship with BMI and weight was less strong for the perioral DSM ( $0.573\leq R^2\leq 0.664$ ), and weak for the perinasal DSM ( $0.305\leq R^2\leq 0.389$ ).

DSM	Sex	% var	BMI	Weight	Height
Whole face	Male	98.1	0.709	0.713	0.365
	Female	98.4	0.701	0.713	0.456
Periorbital	Male	96.8	0.643	0.695	0.396
	Female	97.2	0.636	0.666	0.389
Perinasal	Male	98.3	0.389	0.389	0.370
	Female	98.6	0.305	0.308	0.296
Perioral	Male	98.1	0.664	0.652	0.384
	Female	98.4	0.573	0.587	0.342

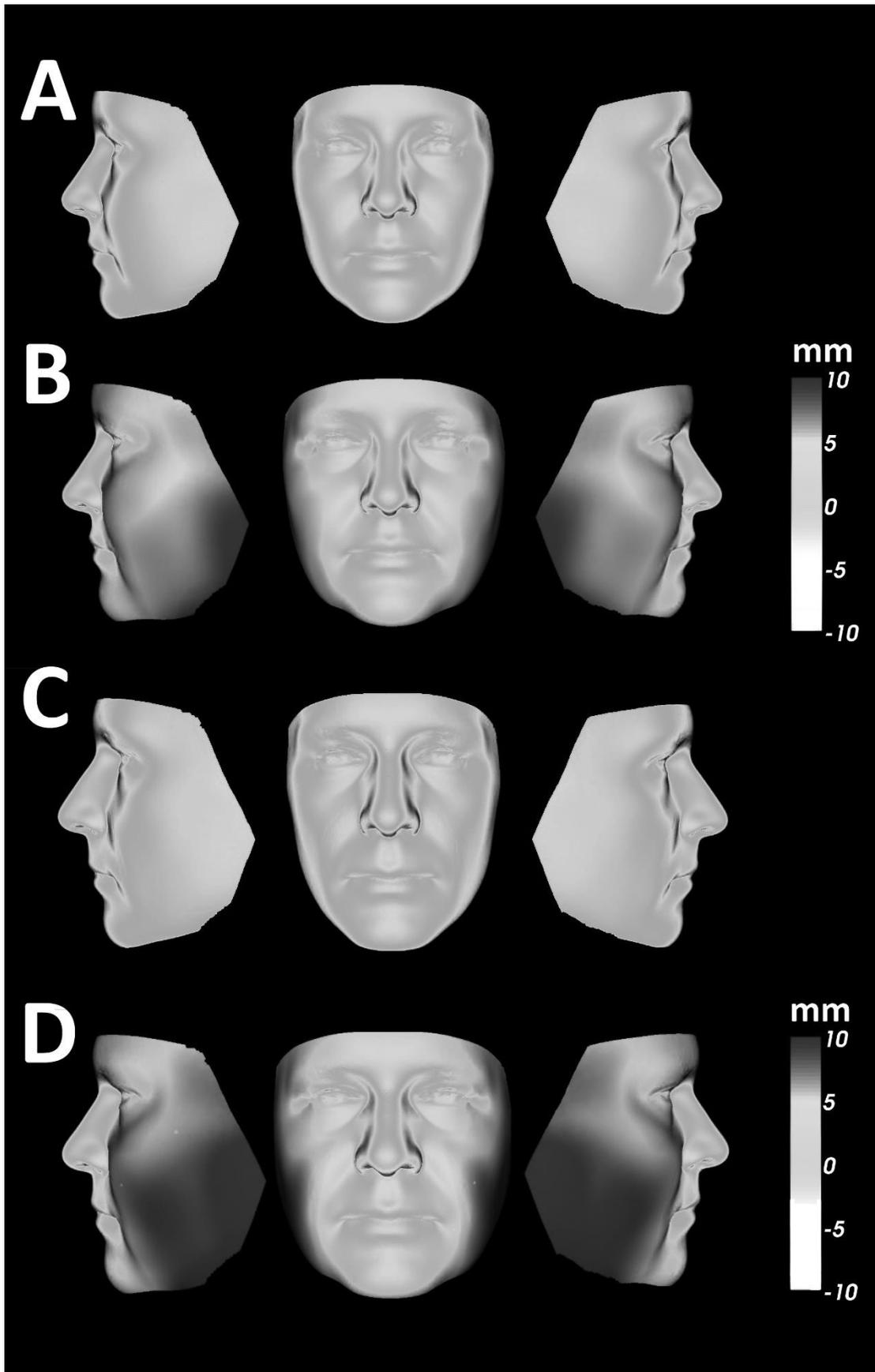
**Table 33** Coefficients of determination ( $R^2$ ) from standard multiple regression analysis are shown. For each of the 4 DSMs, analysis was of 196 men and 236 women. Regression analysis was conducted using a limited number of PCs as variables (39 for men; 47 for women), and the degree of overall shape variation expressed by these modes is shown as a percentage (% var). The strongest association with BMI and body weight is for the whole face DSM. The periorbital and perioral DSMs show a weaker association, and the perinasal DSM is the least associated with BMI and weight. Height showed a weaker association with all DSMs.

#### 6.4.3 GRAPHICAL REPRESENTATION OF FACIAL CHANGE WITH BMI

I used the whole face DSM to see which areas of the face change with BMI graphically.

To avoid bias from individual face characteristics, minimum groups of 30 face images

were chosen at each extreme of BMI values. The mean underweight female face (BMI 19.5; n=30) was compared with the mean obese female face (BMI 41.2; n=30), and was associated with change in facial width. The mean underweight male face (BMI 20.6; n=30) was next compared to the mean obese male face (BMI 36.0; n=30), and similar facial shape changes were seen, but more markedly so. The changes are easier to detect on a computer monitor in colour but images are shown below to aid understanding (Figure 57).



**Figure 57 (A) The mean face of the most underweight women (mean BMI=19.5; n=30) is shown from the whole face DSM in profile and frontal views. (B) The mean face of the most overweight women (mean BMI=41.2; n=30) is shown under this. On screen, it is colour-coded on a red-green-blue axis that represents inward-null-outward displacement, respectively, along the surface normal of the mean underweight female face. Above, it is shown in grayscale with red-green-blue replaced by white-grey-black, respectively. (C and D) The corresponding images are shown for the most underweight men (mean BMI=20.6; n=30) and the most overweight men (mean BMI=36.0; n=30). All images are to the same scale, and the mean male faces are larger than for females. In both sexes, no inward displacement is seen with an increase in BMI, as shown by the lack of white. The midline structures of the face, including the orbits, the nose and the lips show little change with BMI (light grey areas in B and D). In contrast, there is marked outward soft tissue displacement on both sides of the face of 5-10mm, which is more pronounced inferiorly (dark grey areas). There is also outward displacement around the lateral half of the orbits and over the maxilla.**

#### 6.4.4 FACIAL SHAPE CHANGE WITH HEIGHT

I explored whether facial shape could predict height. Standard multiple regression showed that facial shape was less useful for prediction of height than for BMI or weight, with coefficients of determination ranging from 0.296 to 0.456 (Table 33). In women, the whole face was the most useful predictor of height ( $R^2=0.456$ ), whereas in men, the periorbital region ( $R^2=0.396$ ) or perioral region ( $R^2=0.384$ ) were the most useful.

#### 6.4.5 SENSITIVITY ANALYSIS FOR AGE

Facial shape may be affected by a number of factors other than BMI and I conducted a further regression analysis after attempting to control for them. Ageing affects the face, and atrophic changes are known to occur particularly from 50 years onwards (Albert et al., 2007). Facial muscle and fat atrophy would be expected to affect calculations of BMI. In earlier studies, this change from ageing was not a significant issue because all subject faces were compared to age-matched control faces. Therefore, here I excluded all participants above 50 years of age (n=59 for males; n=55 for females). These exclusions mean there were fewer male subjects (119) and fewer female subjects (163). To maintain the minimum 5:1 ratio of cases to variables, standard multiple regression was reassessed using only the first 23 PCs in 119 men, which resulted in a similar coefficient of determination to before ( $R^2=0.709$  for BMI;  $R^2=0.722$  for weight). In 163

women, I used the first 32 PCs and found a similar coefficient of determination ( $R^2=0.666$  for BMI;  $R^2=0.623$  for weight). The coefficients for height were similar to before ( $R^2=0.349$  in men;  $R^2=0.443$  in women). Therefore, there is no significant effect on the relationship of PCs to BMI, weight or height, when including people over 50 years.

#### 6.4.6 ADJUSTMENT OF FSD FOR BMI

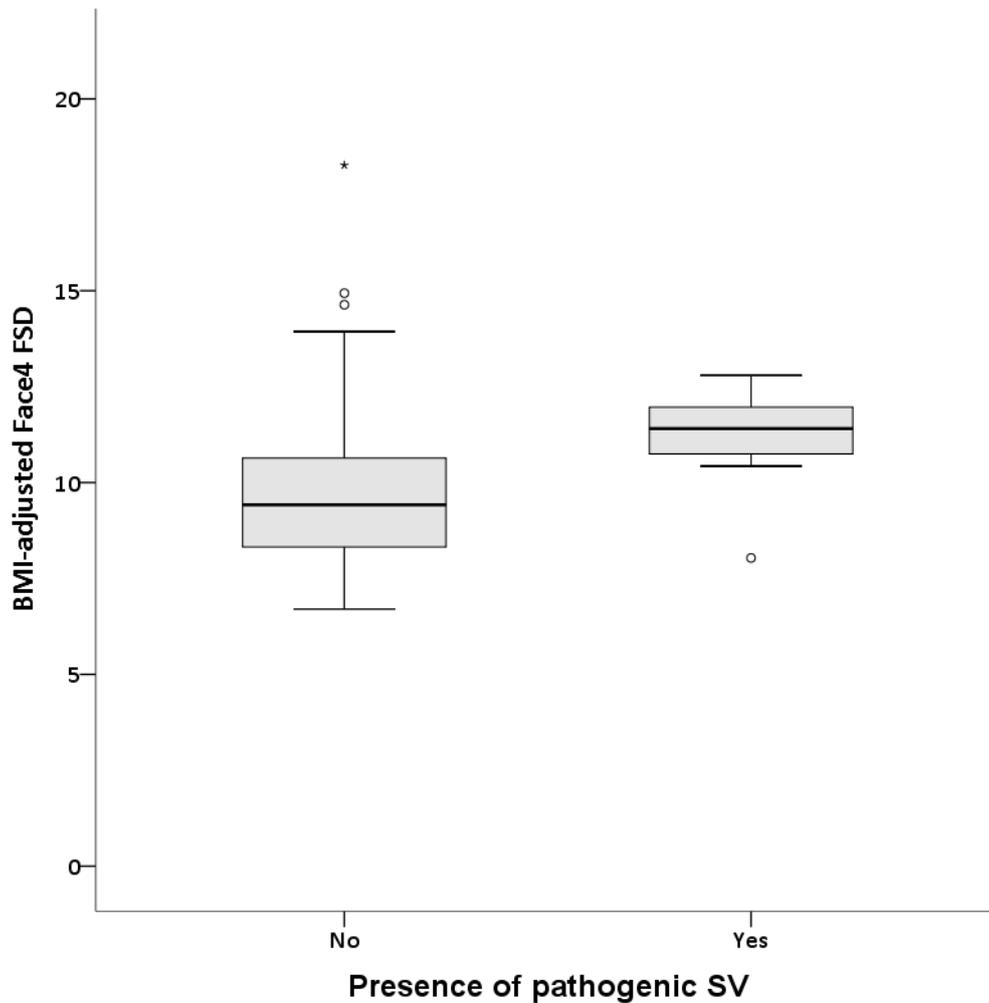
To test the second hypothesis, a DSM of the whole face was created comprising non-genotyped subjects in a training set and genotyped subjects in a validation set, collectively called group 4. The training set included 147 male and 184 female subjects with epilepsy, who had not been genotyped. The validation set contained 98 patients with epilepsy, who had been genotyped. Nine of the 98 patients had SVs. As already shown for group 3, in this new group and DSM, BMI was still significantly associated with each of PC1, PC2 and PC3 in both males and females and linear regression equations were then calculated (Table 34).

		<b>PC1</b>	<b>PC2</b>	<b>PC3</b>
Females	<i>B</i>	-0.029/kgm <sup>-2</sup>	0.048/kgm <sup>-2</sup>	-0.031/kgm <sup>-2</sup>
	<i>R</i> <sup>2</sup>	0.231	0.167	0.087
	Significance	$p=4.8 \times e^{-12}$	$p=8.7 \times e^{-9}$	$p=0.00005$
Males	<i>B</i>	-0.053/kgm <sup>-2</sup>	0.088/kgm <sup>-2</sup>	-0.092/kgm <sup>-2</sup>
	<i>R</i> <sup>2</sup>	0.254	0.231	0.249
	Significance	$p=7.8 \times e^{-11}$	$p=7.1 \times e^{-10}$	$p=1.3 \times e^{-10}$

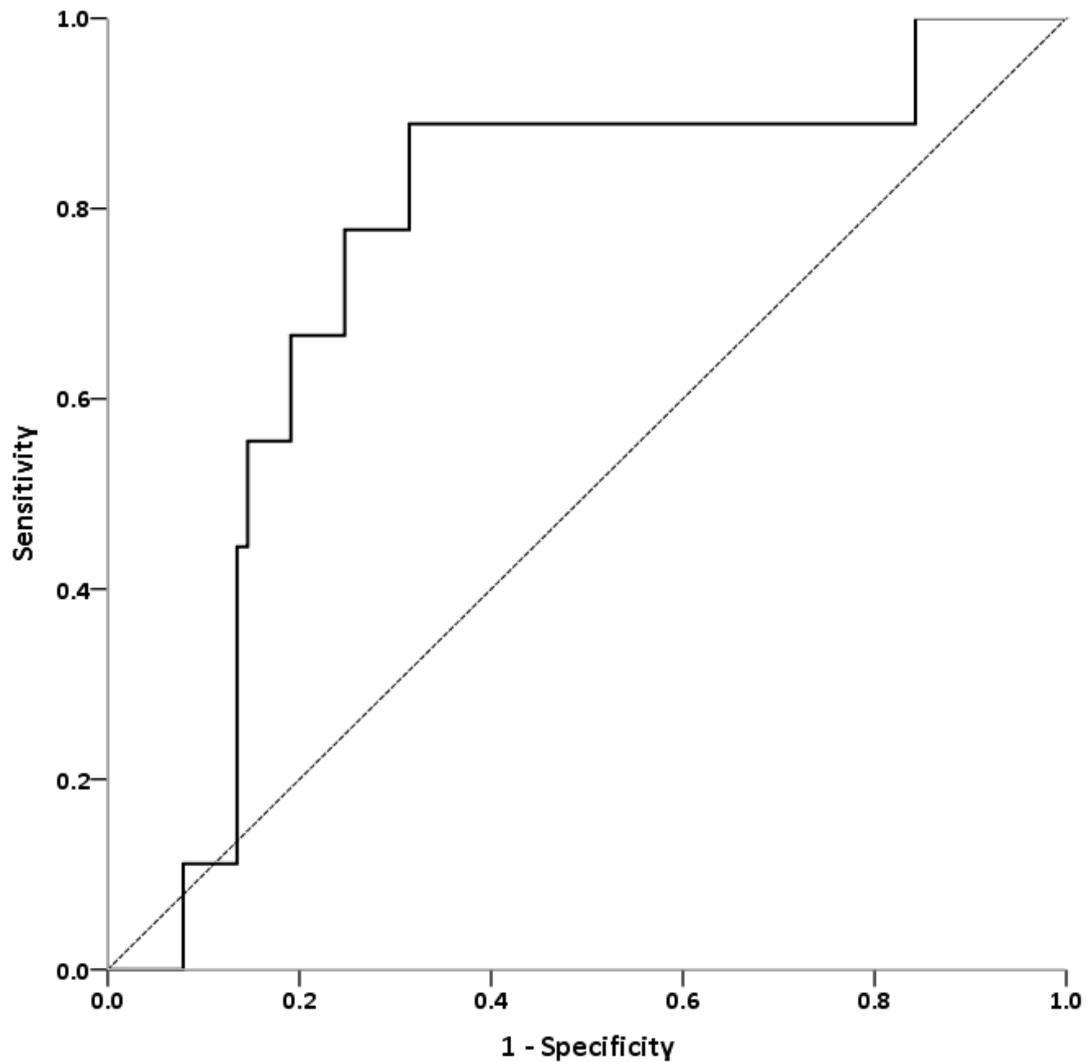
**Table 34 Simple linear regression equations for the first three PCs in male and female subjects in a training subgroup of group 4, with the coefficients here used to adjust the first three PCs for the effect of BMI using BMI=20kg<sup>m-2</sup>**

BMI-adjusted FSD was still significantly greater in people of known BMI with SVs than in people without (11.4 vs 9.42;  $p=0.013$ ;  $n=98$ ; Figure 58). A receiver operating characteristic (ROC) curve showed that a BMI-adjusted FSD cut-off of 10.69 has a sensitivity of 77.8% and specificity of 75.3% in detecting people with SVs (Figure 59). This is more accurate than using FSD values that were not adjusted for BMI (see Chapter 4, Figure 40), which is confirmed by an area under the curve of 0.75. This area

under the curve compares to areas under the curve of 0.61, 0.64 and 0.69 using the unadjusted perinasal region FSD, periorbital region FSD and whole face FSD in all genotyped patients. Therefore, BMI-adjusted FSD is better able to discriminate between people with SVs and people without SVs within my cohort, than unadjusted FSD values.



**Figure 58** FSD of the whole face, after adjustment for BMI differences, remains greater in people with SVs (n=9) than in people without SVs (n=91). The analysis was only performed in people who had been genotyped to look for pathogenic SVs and had had BMI measured.



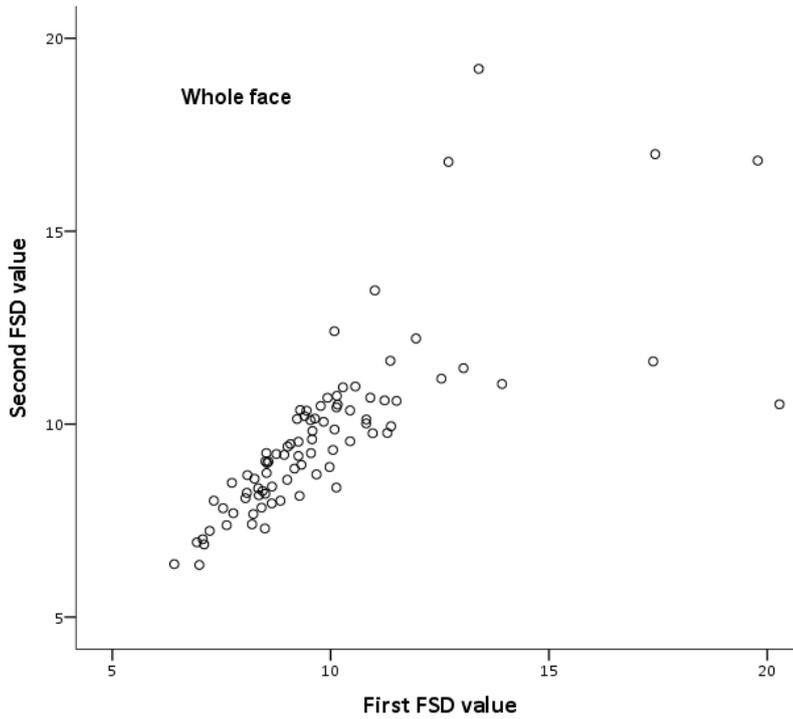
**Figure 59** A receiver operating characteristic (ROC) curve of BMI-adjusted FSD as a method of detecting people with SVs ( $n=9$ ) in a total sample of 98 genotyped subjects with epilepsy. The area under the curve is 0.753 and an optimal BMI-adjusted FSD cut-off value of 10.69 gives a sensitivity of 77.8% and specificity of 75.3%.

#### 6.4.7 LONGITUDINAL STUDY OF CHANGES IN FSD WITH AGE

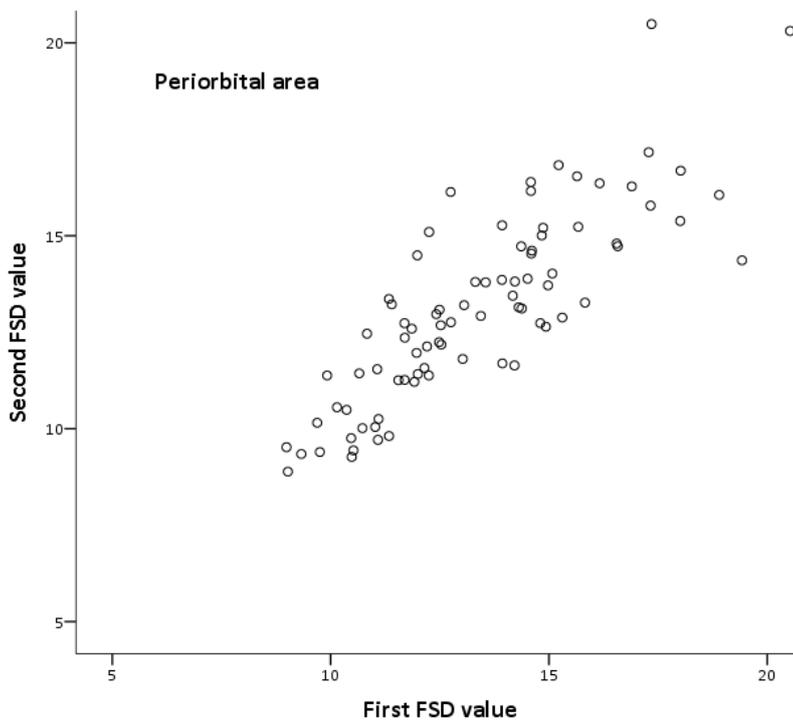
There were 105 people with epilepsy who underwent repeat stereophotogrammetry during a subsequent clinical encounter. Of these, 19 people were excluded because the subject was too old to have suitable age-matched controls (>52 years for men; >54 years for women). Thus, 86 people were included with at least two images taken. The mean age was 37.8 years and 51 were female subjects. The time interval varied from 2 weeks to 2.6 years (mean 1.3 years; median 1.2 years). The mean BMI was 28.1 in the set of first images of subjects and 28.6 in the set of second images of subjects, which was not significantly different (Paired samples t-test;  $p=0.16$ ). Overall, there was no

difference in Face5 FSD (9.37 vs 9.38;  $p=0.35$ ;  $n=86$ ), Eyes5 FSD (13.2 vs 13.1;  $p=0.085$ ;  $n=86$ ) or Mouth5 FSD (7.06 vs 6.95;  $p=0.15$ ;  $n=86$ ). For the perinasal DSM, Nose5 FSD was significantly smaller from the first image to the second image (8.84 vs 8.66;  $p=0.008$ ;  $n=86$ ). This suggests that FSD of the whole face, the periorbital region or the perioral region, can be a robust measurement of face shape that does not significantly vary in a period of up to two years in an individual.

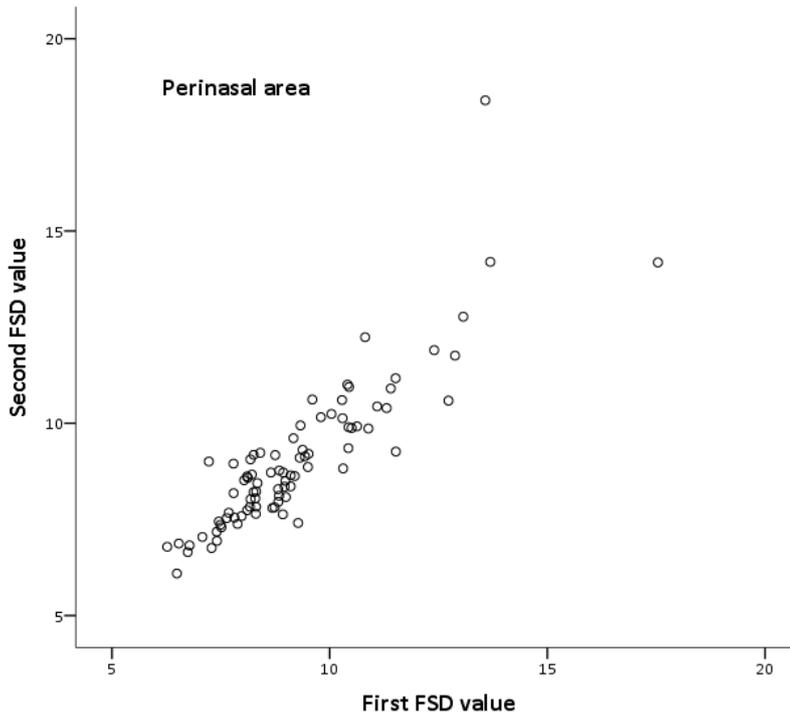
Looking at individual images, one could expect that FSD values should be identical between the initial and repeat images, and a study of correlation can confirm this. Pearson's correlation coefficient suggests that a strong positive correlation between initial and repeat FSD values exists in the whole face DSM ( $r=0.75$ ;  $n=86$ ; Figure 60), perinasal region ( $r=0.87$ ; Figure 62) and perioral region ( $r=0.74$ ; Figure 63). In the periorbital region, correlation is weaker ( $r=0.63$ ; Figure 61). A scatter plot for each DSM shows that correlation appears to be weakest in those people with the highest FSD values. In other words, for some people with high FSD value at one time point, the FSD could be substantially lower at the other time point. This may mean that in some people, FSD is high due to technical error during image capture, such as poor positioning, given that my earlier reproducibility study has shown that such factors can raise FSD in the same subject. Alternatively, it may also mean FSD is more likely to change during a long time interval and I explored this next.



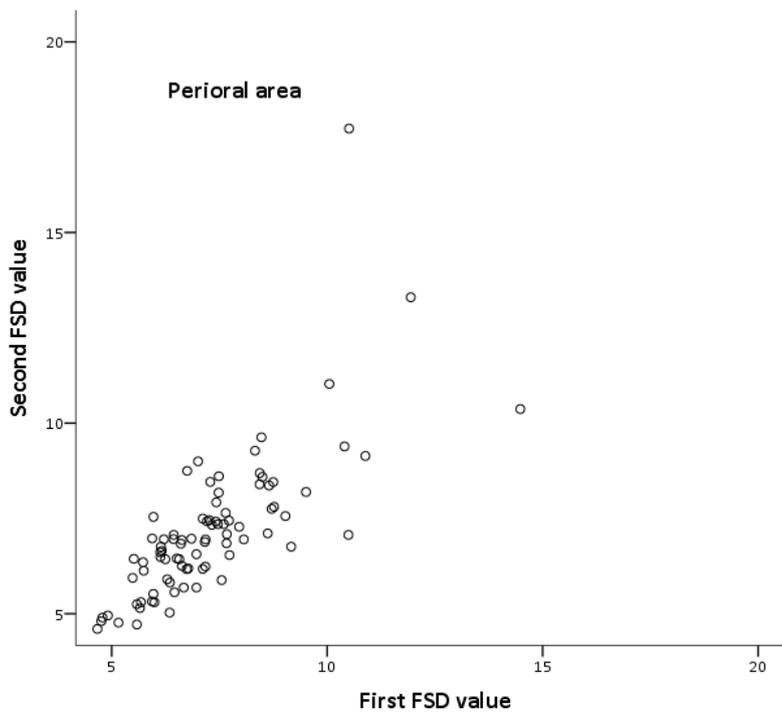
**Figure 60** Correlation between initial FSD and repeat FSD of the whole face for people who underwent serial imaging. Overall, there is strong positive correlation ( $r=0.75$ ;  $n=86$ ).



**Figure 61** Correlation between initial FSD and repeat FSD of the periorbital region for people who underwent serial imaging. Overall, there is moderate positive correlation ( $r=0.63$ ;  $n=86$ ).

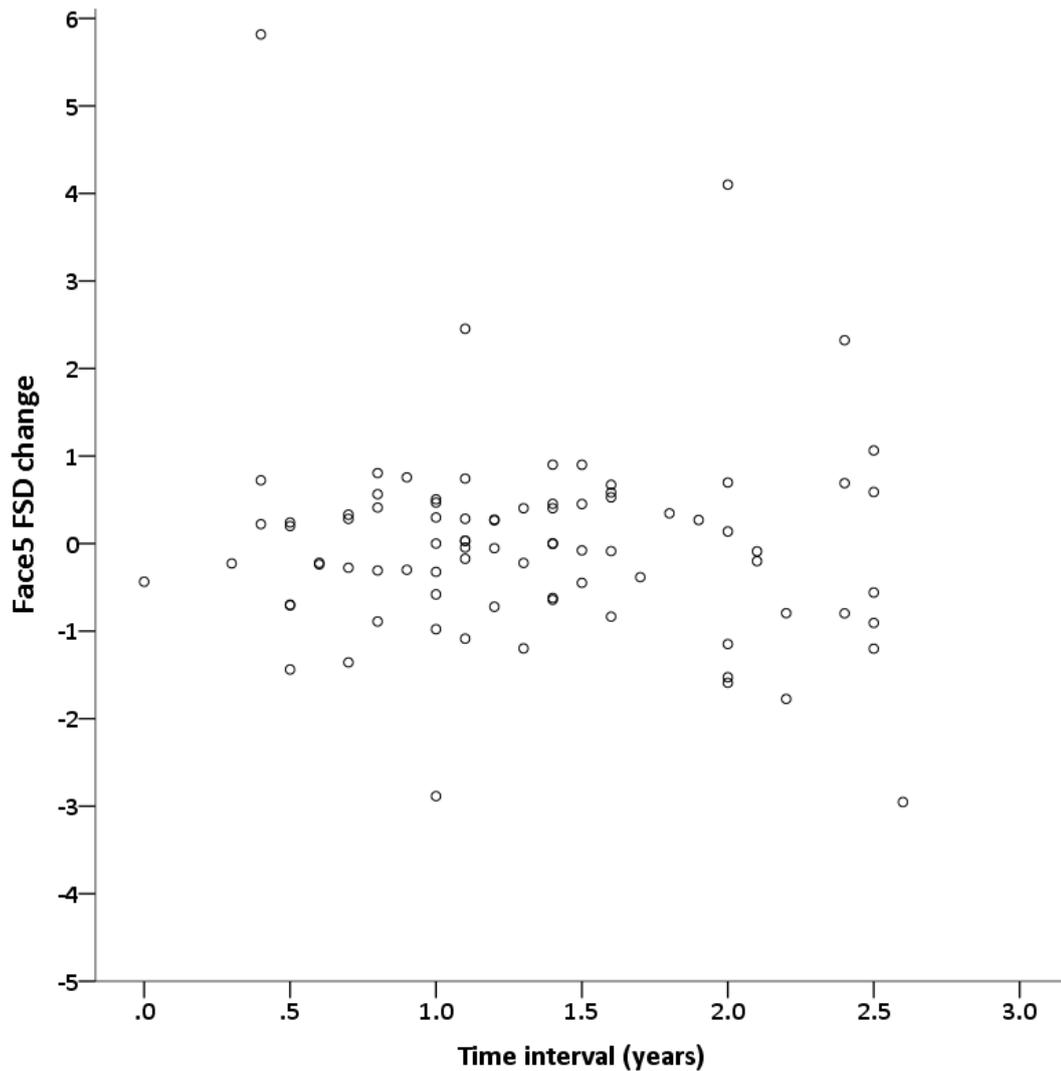


**Figure 62** Correlation between initial FSD and repeat FSD of the perinasal region for people who underwent serial imaging. Overall, there is strong positive correlation ( $r=0.87$ ;  $n=86$ ).



**Figure 63** Correlation between initial FSD and repeat FSD of the perioral region for people who underwent serial imaging. Overall, there is strong positive correlation ( $r=0.74$ ;  $n=86$ ).

The change in Face5 FSD was assessed against the time interval between images and there was no correlation, suggesting that there is no systematic error in measurement, although a scatter plot does show there are a number of outliers with repeated FSD (Figure 64).



**Figure 64** Scatter plot of change in Face5 FSD with the time interval between the first image and second image in 86 subjects. Overall the median difference is not significantly different between the first and second image (9.37 vs 9.38;  $p=0.35$ ;  $n=86$ ). It can be seen though that there are some outliers with 8 of 86 subjects having a change in Face5 FSD of greater than 2. Outliers were seen at any time interval between 0.5 – 2.7 years and this did not increase with time.

Eight of 86 subjects have a change in Face5 FSD of greater than two. Visual inspection of the extreme outliers with the greatest difference in FSD between the two images

shows that there are visible discrepancies that explain the change in FSD. The FSD for regions of the face are very useful in determining the source of the discrepancy.

For the three faces with the biggest change in Face5 FSD, the FSD values are shown for all DSMs (Table 35).

	<b>Change in FSD between both images</b>			
	Face5	Eyes5	Nose5	Mouth5
Subject 1	5.82	11.24	-0.37	0.00
Subject 2	-5.76	-20.47	-2.25	-0.06
Subject 3	-9.76	-0.22	-0.61	-0.82

**Table 35** *The change in FSD in all DSMs between the first and second image for the three subjects with the greatest change in Face5 FSD. The table shows that FSD in the periorbital DSM (Eyes5), perinasal DSM (Nose5) or the perioral DSM (Mouth5) did not necessarily correlate with changes in Face5 FSD. In subjects 1 and 2, the change in Face5 FSD appears to be due to a difference in the upper face and periorbital region with minimal difference in the perioral region. In subject 3, the difference in Face5 FSD is due to a difference in the outer border of the face which is not captured by the DSMs of parts of the face.*

In subjects 1 and 2, the change in Face5 FSD appears to be due to a difference in the upper face and periorbital region with minimal difference in the perioral region and a slight difference in the perinasal region. In subject 3, the difference in Face5 FSD is due to a difference in the outer border of the face, which affects Face5 FSD. However, this difference occurs in a part of the face that is not included in the periorbital, perinasal or perioral regions; hence these DSMs of parts of the face are not affected. Inspection of the raw images confirms the expected differences in the face surface (Figure 65).

Subjects 1 and 2 have hair obscuring part of the forehead in the first image, which also affected the underlying face surface. Subject 3 also has hair obscuring the side of the face in the first image, which is less apparent in the second image.

## First image      Second image



***Figure 65 Cropped images of the foreheads of three subjects with the largest changes in Face5 FSD between the first and second image. Hair obscures the forehead and periorbital region in subject 1 (top row) and subject 2 (middle row) in the first image but not the second image. In subject 3, hair covers the side of the face in the first image but not the second image, although this is not easily seen in the cropped images.***

Note that in the earlier study with pathogenic SVs, coverage of the face was assessed on visual inspection after image capture by a physician blinded to clinical details and a sensitivity analysis was conducted excluding those with visible acquired deformities on the face.

### 6.5 DISCUSSION

I used 3D stereophotogrammetry and dense surface modelling to analyse facial shape in a UK population of people with epilepsy and their relatives. The first part of this study demonstrates that weight and BMI do significantly affect face shape, as expected. Face shape analysis using PCs can determine 60-70% of the variance in BMI and weight. More importantly, BMI and weight affect PCs used in face shape analysis in a predictable manner. With increasing BMI, the predominant change seen was an

increase in soft tissue width over the sides of the face. This area is only included in the whole face DSM and this DSM shows stronger correlation with BMI. The perinasal DSM has the weakest association with weight and BMI, suggesting that there is little soft tissue change around the nose with an increase in weight. This is confirmed on the colour-coded facial images. Perinasal DSM therefore may not be sensitive in detecting syndromes associated with change in weight or BMI. On the other hand, perinasal DSM should also be the least affected by obese or underweight subjects who may confound FSD in other DSMs. For example, Smith-Magenis syndrome is associated with facial dysmorphism, including the nose, and obesity. An obese person would be expected to have increased FSD in all DSMs, but least of all in the perinasal DSM. An obese person with Smith-Magenis syndrome would be expected to have increases in PCs for all DSMs due partly to increased BMI, but least affecting the perinasal DSM. Thus, the perinasal DSM may be best able to detect this syndrome in a group of people of different weight and BMI. The periorbital and perioral DSMs are more strongly associated with weight and BMI, with coefficients of determination over 0.57. This may be because these DSMs include more of the facial width and so are sensitive to the increases in width that occur with increasing weight.

There is a weaker relationship between facial shape and height. Forensic studies in other ethnicities using skull measurements have found coefficients of determination of between 0.10 and 0.36 between height and craniofacial measurements (Chiba and Terazawa, 1998; Pelin et al., 2010). I found that my methods give coefficients from 0.30 to 0.46 using any one of the DSMs. This suggests that using DSMs of facial regions instead of skull measurements does not detract from the accuracy of estimation of height. It is another application in which stereophotogrammetry and dense surface modelling are equivalent or superior to traditional anthropometry, although this is of limited clinical relevance.

Another study in a UK population used a photonic scanner to look at body shape and weight in over 9000 people (Wells et al., 2007). This involved subjects undressing and maintaining a posture for 6 seconds for image capture. The authors found that correlation coefficients ranged from 0.67 to 0.92 between weight and selected body measurements, such as hip circumference. This would suggest that  $R^2$  ranges between 0.45 and 0.85. My study found  $R^2$  up to 0.71 using DSMs, in a smaller group of subjects, using a less 'invasive' and faster method. This suggests that more detailed techniques using measurements in the whole body provide no added benefit in estimating weight than using face shape alone.

BMI and weight themselves do not account for the effect of distribution of fat, and measures such as waist circumference and waist-height ratio are thought to be more useful clinically to assess the health risks of obesity (Ashwell et al., 2012). Patterns of craniofacial fat deposition may predict particular obesity-related disorders or mortality, and could be detected using stereophotogrammetry and DSMs. For example, patients with OSA, which is strongly linked to obesity, have a distinct craniofacial profile (Banabilh et al., 2009; Lee et al., 2010). Currently, methods of describing these craniofacial changes use cephalometry, with radiation exposure, and/or cumbersome intra-oral measurements (Lam et al., 2005).

Many genetic obesity syndromes are associated with facial dysmorphism and it has been suggested that others are still unrecognised (Farooqi and O'Rahilly, 2005). One example is the novel finding of 16p11 deletions in 0.7% of the morbidly obese population (Walters et al., 2010). DSMs have already been used to identify dysmorphic syndromes, and may also aid identification of atypical face shape in obese subjects from comparison of facial shape with that expected for an average weight, height and age. DSMs are well-suited for facial shape analysis in obesity: conventional anthropometry relies on easily identifiable landmarks on the face, which are mostly located around the

eyes, nose, mouth, ears or bony landmarks, but I show that obesity mainly affects the lateral aspect of the face, which has few such landmarks.

I also show that conversely, in cases where obesity may hide atypical facial features, FSD may be adjusted to account for BMI. This correction study was performed in a different group, group 4, to that exploring the relationship of BMI to face shape initially. The reason for this was that the first hypothesis was to see how much of face shape variation was associated with weight and BMI. To be more accurate, the largest sample size possible was used, including relatives without epilepsy. Exclusion criteria were only strict in avoiding potential serious confounding factors, such as those with any genetic syndrome and those with any known facial injury. In the study of the second hypothesis, I was demonstrating how FSD could be corrected for weight and BMI. To do this, the study sample used people with epilepsy only, for whom I have used FSD already. This study then uses my method of BMI adjustment to see how it affects an earlier analysis I performed in people with and without SVs. In Section 4.4.12, I acknowledged that BMI may be a confounder in the analysis of people with and without SVs, but had no means of adjusting FSD for this other than to exclude people with BMI outside the accepted normal range. Here I was able to show that FSD, adjusted for BMI, is still able to discriminate between people with pathogenic SVs and people without. The effectiveness of this method is still to be assessed in larger samples. The first step is to repeat validation in a larger sample to see if BMI-adjusted FSD is superior to FSD in detecting SVs. Adjustment for BMI has to be performed with care, as it represents additional recalculation that may lose some of the information from PCs and the unadjusted FSD value. Indeed, for novel syndromes or novel SVs, it may not be known whether BMI or weight are affected and adjusting FSD for BMI would make it impossible to establish this.

A limitation of BMI analysis is that the BMI of reference subjects is not known. It has been assumed that using the mean face of 30 reference subjects is sufficient to avoid changes in BMI between the mean matched faces. It would be prudent to check this in future studies, perhaps by recruitment of different reference subjects with BMI measurements. This is further discussed in Chapter 9. Another clear study limitation was that most participants had epilepsy. People with epilepsy tend to have a greater BMI than those without, partly because of a more sedentary lifestyle (Ben-Menachem, 2007). Epilepsy itself is not known to affect facial shape, but may occasionally be associated with underlying genetic or developmental syndromes that do cause dysmorphism. All known such cases were not included. I did not assess the effect of anti-epileptic drugs. These drugs may cause weight gain or weight loss (Ben-Menachem, 2007).

For the ageing study, the perinasal DSM showed a significant decrease in FSD over time but the other DSMs do not. The decrease is small (FSD 8.84 to 8.66). For the other parts of the face, there is no significant change. There may be two reasons for this. Firstly, some of those with epilepsy in this group may develop a more typical face shape (especially the nose) with age. This is known to occur in certain syndromes, such as 22q11 deletion syndrome (Hammond et al., 2005). Secondly, an atypical nose shape in the initial image may be from poorly controlled seizures and trauma. In Chapter 4, it was seen that nearly 5% of patients with epilepsy had suffered from visible deformity of the nose due to a suspected fracture.

Excluding the perinasal DSM though, it appears that FSD is a measure that is not affected by age within a time interval of 2 years. The time interval is too short for a detailed study of how FSD may change with age. Given that FSD is a measurement that is matched for age, one would not expect FSD to change significantly in control subjects, as their comparators should show the same ageing changes. Factors that could cause

change include weight change, facial injury, underlying genetic syndromes and technical errors during image capture. Some dysmorphic syndromes become less prominent with age and it would be interesting to see if FSD is less able to detect such syndromes with age too. Other characteristics in the repeated images of subjects such as individual PCs or inter-landmark distances were not assessed. Previous studies have shown that the face changes with age, but these changes are of the order of a few millimetres over decades. Stereophotogrammetry would probably be able to detect these, but one needs a much longer time interval. Indeed, to my knowledge, no longitudinal study on face shape during ageing has been conducted using stereophotogrammetry because of the relative infancy of this technique. Nonetheless, my findings in this ageing study are useful as it corroborates my earlier reproducibility studies. FSD is calculated by matching each subject's face to the mean face of the 30 closest reference images by age and sex. When FSD is calculated after the median time interval of 1.2 years in this study, some of the underlying matched reference images are different as the subject is older. With an increasing time interval, the proportion of matched reference images that are different becomes greater and hence the mean face generated is likely to differ more. However, I did not see any association between change in FSD and the time interval between images in this small sample over a relatively short time interval. This means that change in the underlying reference images used does not appear to make a measurable difference to FSD. It suggests that the method of using 30 matched reference images to calculate FSD is sound.

An important finding in the ageing study is that FSD correlates more strongly in repeated measurements over time when FSD is smaller. A minority of people with a high FSD were not found to have a high FSD when measured at a different time interval (either in the first or the second image). Visual inspection of the face images and the other FSD values is useful in this situation to clarify FSD differences due to inadequate image capture. It confirms the importance of ensuring the face surface is fully covered

by stereophotogrammetry. If this is not possible, then only a DSM of part of the face should be used – for example, the perinasal and perioral DSMs were robust to the changes in examples shown earlier in this chapter. The first images in most subjects in the longitudinal ageing study were taken early during recruitment. By the time of second images, my image capture technique and protocol had been refined and this may explain part of the discrepancy. For example, a hairband was used to shift long hair off the face surface.

A major limitation of the ageing study is that I only had one time interval between images and that interval was less than 3 years in all cases. A longer longitudinal ageing study would show if age confounds face shape analysis despite the use of age-matched controls. Also there was no clinical information on drug use or history of trauma between the images.

## 6.6 CONCLUSION

In conclusion, I have shown that changes in facial shape are clearly associated with BMI and weight and illustrated the location and extent of these changes. I also show that 3D stereophotogrammetry and dense surface modelling are useful methods for analysis of facial shape changes. The effect of BMI on facial shape can be reduced using recalculating FSD after adjusting for BMI. Reanalysing a study presented in Chapter 4, it is shown that adjustment for BMI does not appear to reduce the usefulness of FSD.

I have also conducted the first analysis of longitudinal changes in DSMs of the face and parts of the face showing that FSD does not change significantly over a period of up to 2.5 years. However, it is important to acquire face images using correct technique and then visually inspect the image to ensure quality.

## CHAPTER 7: FACE SYMMETRY IN DEVELOPMENTAL BRAIN LESIONS

### 7.1 INTRODUCTION

Human face shape is known to be dependent on underlying brain development (Marcucio et al., 2015). This has long been recognised in conditions such as holoprosencephaly, microcephaly or lissencephaly, which are associated with facial malformations (Demyer et al., 1964; Winter, 1996). One explanation for this association is the physical proximity of the anterior neural tube and forebrain to the facial prominences. In mice, different rates of brain growth are known to affect the size and shape of the embryonic face (Marcucio et al., 2011). The brain also produces a number of molecular signals that are important for facial development, such as SHH or BMP proteins, and the interruption of these signalling pathways cause facial malformations (Marcucio et al., 2005). A further reason is that facial structures primarily derive from neural crest cells, which in turn form from the dorsal neural tube (Cordero et al., 2011).

Abnormal brain development may also manifest as epilepsy. This can occur from both genetic causes, such as in tuberous sclerosis, or from acquired causes, such as in hippocampal sclerosis (Malmgren and Thom, 2012). However, there is often overlap. For example, people with febrile seizures, mesial temporal lobe epilepsy and hippocampal sclerosis were more likely to have a common variant in the *SCN1A* gene (Kasperavičiūtė et al., 2013). With the advent of higher resolution magnetic resonance imaging (MRI) of the brain, more subtle disorders of neuronal migration and development have also been discovered to be a cause of epilepsy (Aronica and Crino, 2014; Raymond et al., 1995). These malformations of cortical development can be divided into a number of types, including band heterotopia, dysembryoplastic neuroepithelial tumours, focal cortical dysplasia and periventricular heterotopia (Sisodiya, 2004). For each of these, genetic causes are now also recognised.

Concurrent with advances in brain imaging, analysis of face shape is now possible in much greater detail with stereophotogrammetry and dense surface modelling. Facial asymmetry has not been studied in people with epilepsy to my knowledge. Most studies have been performed in orthodontics. A review article in 2015 states that visible facial asymmetry has a prevalence of 12-74% in the general population with no difference seen in different ethnicities (Thiesen et al., 2015). Most of the studies were those predominantly assessing deviation of the mandible and maxilla from the midline. The method of face shape analysis, such as anthropometry or cephalometry, and the strictness of the criteria used, determine the prevalence of asymmetry. For example, 85% prevalence of facial asymmetry was found by a group which defined asymmetry as >2mm deviation from the midline for any midline facial bone landmark. Dense surface modelling has been successfully used before to show that there is increased facial asymmetry in boys with autism (Hammond et al., 2008), a neurodevelopmental disorder, which may be due to a range of genetic factors (de la Torre-Ubieta et al., 2016). Developmental brain lesions, also called malformations of cortical development, could conceivably be the result of abnormal forebrain development and thus also affect ipsilateral (or contralateral) facial development if unilateral.

## 7.2 HYPOTHESES

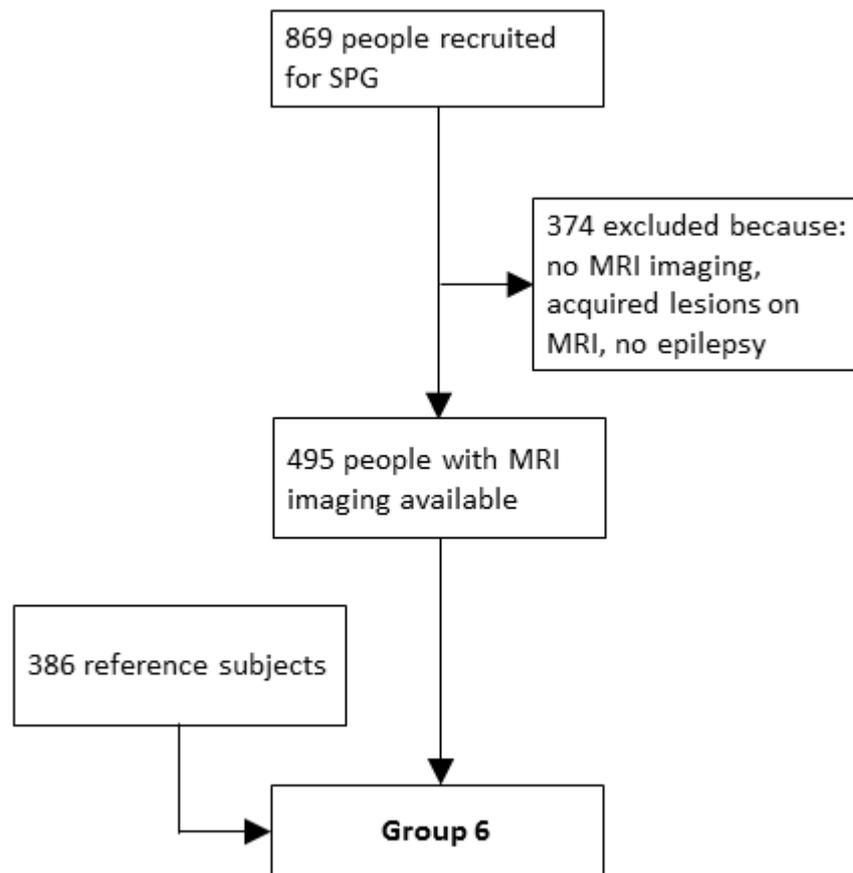
I hypothesised that facial asymmetry can be quantified using dense surface models generated from stereophotogrammetry and that facial asymmetry was increased in people with epilepsy and underlying malformations of cortical development compared to that in people with epilepsy and no such malformations.

## 7.3 METHODS

### 7.3.1 RECRUITMENT, IMAGE CAPTURE AND FACE SHAPE ANALYSIS

Participants were recruited as described for group 6 in Chapter 2 (Figure 66). Out of 869 people recruited in total, 374 were excluded due to one of the following reasons:

subject not known to have epilepsy, no MRI images or reports available for the subject or the presence of an acquired lesion, an undefined lesion or multiple lesions on the MRI scan. Undefined lesions included any lesion that could not be diagnosed and also includes those MRI scans deemed to have significant movement artefact and insufficient quality to exclude a lesion by the reporting radiologist. Acquired lesions comprised stroke, traumatic injury, cavernomas, brain atrophy, facial surgery, intracerebral cysts and tumours. Any person with multiple lesions, such as hippocampal sclerosis with periventricular heterotopia, was also excluded.



**Figure 66 Recruitment of subjects into group 6. Overall, 495 people with epilepsy and MRI results available were included in the study, along with 386 reference subjects used for matching to help calculate FSD. SPG = stereophotogrammetry**

Image capture, was performed as in Chapter 2. All face images were assessed by one of two operators (Vamsi Chinthapalli or Jan Novy), blinded to the MRI findings at the time. Any acquired image deformity was recorded for each image as well as the location of the deformity. Acquired deformity comprised facial hair, piercings, scars or raised skin lesions. Those images with the mouth open were noted. The location of any deformity was used to assess which dense surface models (DSMs) should be excluded in a sensitivity analysis. For example, a patient with a suspected previous nasal fracture would typically be excluded from the Face6 and Nose6 DSMs but not from Eyes6.

### 7.3.2 DENSE SURFACE MODELLING

Dense surface modelling was performed as for all previous DSMs with the additional step below of mirror image creation.

Mirror images were created after landmarking using an automated procedure in the DSMExplorer software program. All paired landmarks on a particular subject's image are reassigned to their opposite landmark (e.g. left exocanthion is labelled as 'right' exocanthion). Before the thin plate spline step, when a face surface is deformed and aligned with a base surface, the relabelling means that a face surface is reflected in the midline to allow the landmarks to match up with the landmarks on the base surface. A dense correspondence is created as usual for this mirror image with subsequent principal component analysis. Both unreflected and mirror images are all included in the resulting DSM. Note that manual landmarking is only performed once for a face image and the mirror image uses the same landmarks, which helps reduce error.

DSMs of the whole face (Face6), periorbital region (Eyes6), perinasal region (Nose6) and perioral region (Mouth6) were created in which 99% of the shape variation was explained using 97, 239, 79 and 92 PCs, respectively.

### 7.3.3 REFLECTED FACE SHAPE DIFFERENCE

After the DSM was created including these mirror images for all subjects, I calculated the distance between each face and its mirror image in terms of the square root of the sum of squared differences of every principal component. This is different to FSD, in which the distance between a face and its *age-matched, sex-matched mean reference face* was calculated. Algebraically, the formula is:

$$\text{FSD}(x,r) = \sqrt{\sum_{i=1}^n (x_i - r_i)^2} \text{ for } x \in M$$

where  $i$  indexes the  $n$  principal components ( $PC1, PC2, PC3$ , etc.) capturing 99% shape variation in dense surface model  $M$  and  $x$  is an arbitrary subject's face which is compared to  $r$ , its reflected face surface image.

This formula has been derived by me alone for this study and the distance was termed the 'reflected face shape difference'. A similar measure has been used before to detect facial asymmetry in autism spectrum disorders (Hammond et al., 2008). Like FSD, reflected FSD only measures the size of any difference in the face images.

One can see that for reflected FSD, control subjects are not necessary as a subject's face is compared only with its reflection.

### 7.3.4 CLASSIFICATION OF DEVELOPMENTAL BRAIN LESIONS

Only those with available brain MRI reports by a neuroradiologist, clinical documentation of MRI findings, or imaging review by an experienced neurologist were included in the study. Based on these findings, subjects were categorised according to the presence or absence of a developmental brain lesion and the side of any lesion. For the purposes of this study, developmental brain lesions were defined as those visible on MRI and comprised: band heterotopia (BH), dysembryoplastic neuroepithelial tumour (DNT), focal cortical dysplasia (FCD), hippocampal sclerosis (HS), periventricular heterotopia (PVH), tuberous sclerosis (TS) or an undefined/other

malformation of cortical development (MCD). Tuberous sclerosis was defined either as the most likely radiological diagnosis based on the presence of cortical tubers or as otherwise clinically confirmed tuberous sclerosis, such as from histology or genetics, with an abnormal MRI scan. I appreciate that the cause of hippocampal sclerosis is likely to be complex, comprising genetic, developmental and acquired factors, including seizure severity (Walker, 2015), but for the purposes of this analysis, it has been included as a developmental brain lesion (labelled as 'HS'). The malformations in the MCD group included people with a type of malformation present in less than four others in the group, people with multiple types of defined malformations, or people with an indeterminate malformation based on imaging. They are listed in Section 7.4.1. It is acknowledged that there are more detailed classifications available for such developmental brain lesions based on clinical features, high-resolution imaging, genetic testing and neuropathology (Barkovich et al., 2012). However, there are limitations of these classifications, which include misuse of terminology, such as polymicrogyria, and the limited use of many categories of malformations in clinical practice (Guerrini and Dobyns, 2014). Another limitation is the meaningfulness of more detailed classification, with recent discoveries that supposedly distinct malformations, such as lissencephaly and polymicrogyria-like malformations, may have the same affected gene.

### 7.3.5 STATISTICAL ANALYSIS

The Mann-Whitney U test was used to compare means between groups. The Kruskal-Wallis test was used to compare people without HS, with left-sided HS, with right-sided HS and with bilateral HS. Analysis was conducted using SPSS (versions 19-23, SPSS Inc; Chicago, IL, USA).

## 7.4 RESULTS

### 7.4.1 SUBJECT POPULATION

I recruited 495 participants. Of these, 329 had no lesion and 166 had a developmental brain lesion. Ninety-three people had HS, 26 had FCD, 12 had PVH, 10 had DNTs, 7 had BH, 5 had TS and 13 had MCD. The median age of subjects with a developmental brain lesion was 7.2 years greater than subjects without a lesion and there was a similar sex ratio in both groups (Table 36). However, there were different sex ratios for different lesion types, with the majority of DNT, FCD, TS and MCD lesions occurring in men, in contrast to the other types.

The thirteen people with MCD comprised: Seven with extensive mixed malformations of cortical development involving cortical and white matter (in one case bilateral), two with polymicrogyria, two with a cortical abnormality similar to focal cortical dysplasia but more extensive, one with schizencephaly, and one with bilateral pachygyria and *DCX* gene mutation.

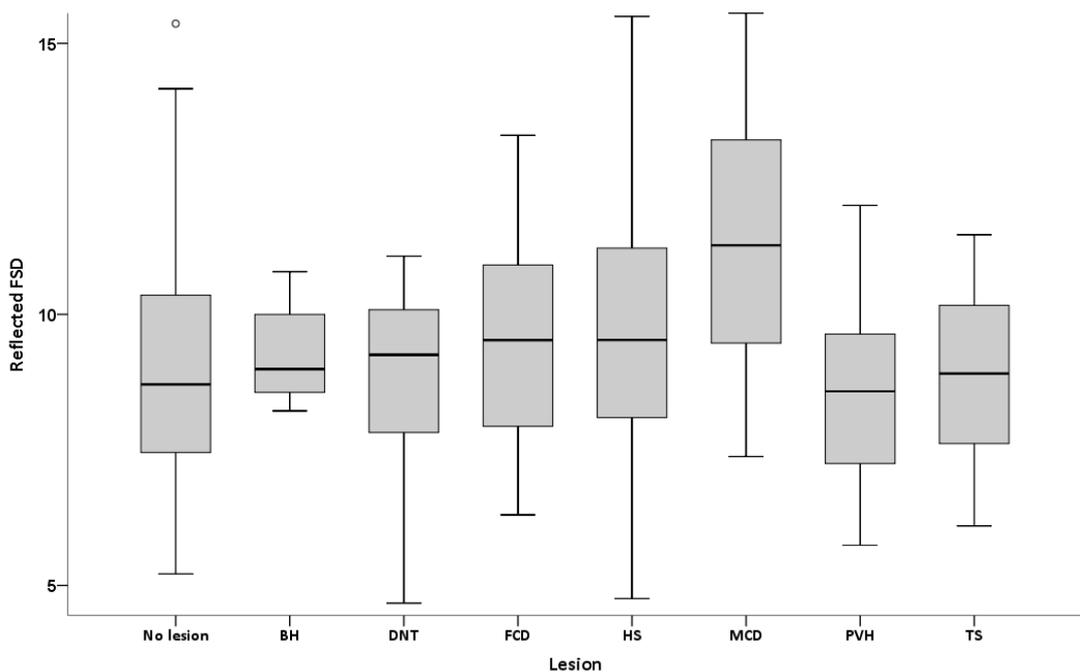
	Structural lesion								
	None	Any	BH	DNT	FCD	HS	PVH	TS	MCD
Number	329	<b>166</b>	7	10	26	93	12	5	13
Unilateral lesion only	-	<b>133</b>	1	10	26	82	4	0	10
Age (median, IQR)	35.0 (17.0)	<b>42.2</b> <b>(14.6)</b>	38.1 (18.9)	43.0 (8.7)	35.6 (17.3)	44.5 (13.7)	38.5 (9.8)	29.6 (10.0)	33.9 (15.2)
Sex (% male)	43.8	<b>44.6</b>	28.6	60.0	69.2	34.4	41.7	60.0	61.5
Median reflected FSD	8.71	<b>9.54</b>	8.99	9.25	9.35	9.53	8.58	8.91	11.27
Significance	N/A	<b>p=0.001</b>	p=0.466	p=0.885	p=0.138	p=0.002	p=0.642	p=0.961	p=0.002
<b>People with no acquired facial deformity</b>									
No acquired deformity	268	<b>135</b>	4	8	17	81	12	5	8
Median reflected FSD of images with no acquired deformity	8.51	<b>9.48</b>	8.82	9.25	9.48	9.52	8.58	8.91	12.53
Significance	N/A	<b>p=0.001</b>	p=0.668	p=0.996	p=0.308	p=0.001	p=0.922	p=0.846	p=0.004

**Table 36** Details for all participants in the study with the number in each group and age and sex characteristics. The median reflected face shape difference (FSD) is shown for the total number in each group as is the significance of the result when compared with the reflected FSD in participants with no malformation of cortical development. The analysis is repeated after exclusion of people with acquired facial deformity as judged from face images only and blind to MRI or clinical information. Overall, developmental brain lesions are associated with significantly greater facial asymmetry as judged by reflected FSD and this is also true for the categories of HS and MCD. There is no change in these findings after exclusion of people with acquired facial deformity. BH = band heterotopia, DNT = dysembryoplastic neuroepithelial tumour, FCD = focal cortical dysplasia, FSD = face shape difference, HS = hippocampal sclerosis, IQR = interquartile range, MCD = other malformation of cortical development, PVH = periventricular heterotopia, TS = tuberous sclerosis

## 7.4.2 REFLECTED FSD IN PEOPLE WITH DEVELOPMENTAL BRAIN LESIONS

Reflected FSD was higher in participants with a developmental brain lesion than in those without (9.54 vs 8.71;  $p=0.001$ ;  $n=495$ ).

To explore this further, I compared individual brain lesions to those without any lesions (Table 36). Both those with bilateral lesions and those with unilateral lesions were included. Subjects with HS had a significantly higher reflected FSD (Figure 67; 9.53;  $p=0.002$ ;  $n=93$ ), as did those with MCD (11.27;  $p=0.002$ ;  $n=13$ ). However, there was no difference in reflected FSD, compared to subjects with no lesion, in subjects with BH (8.99;  $p=0.466$ ;  $n=7$ ), DNT (9.25;  $p=0.885$ ;  $n=10$ ), FCD (9.35;  $p=0.138$ ;  $n=26$ ), PVH (8.58;  $p=0.642$ ;  $n=12$ ) and TS (8.91;  $p=0.961$ ;  $n=5$ ).



**Figure 67** Box plots of the distribution of reflected FSD for each type of developmental brain lesion. MCD had a significantly greater reflected FSD compared to 329 people with no MRI lesions (11.27 vs 8.71;  $p=0.002$ ;  $n=13$ ). HS also had a greater reflected FSD (9.53 vs 8.71;  $p=0.002$ ;  $n=93$ ). People with DNT and FCD had similar reflected FSD values to HS, but they were smaller groups and reflected FSD was not significantly different to people with no MRI lesions. BH = band heterotopia, DNT = dysembryoplastic neuroepithelial tumour, FCD = focal cortical dysplasia, FSD = face shape difference, HS = hippocampal sclerosis, IQR = interquartile range, MCD = other malformation of cortical development, PVH = periventricular heterotopia, TS = tuberous sclerosis

### 7.4.3 REFLECTED FSD IN UNILATERAL AND BILATERAL LESIONS

Thirty-three participants had bilateral developmental lesions seen on MRI. This included the majority of those with BH (6 of 7), PVH (8 of 12) and TS (5 of 5). By contrast, all those with a DNT or FCD had unilateral lesions only. I looked at unilateral and bilateral lesions separately with the hypothesis that unilateral lesions may be associated with greater facial asymmetry. Both those with unilateral lesions (9.48;  $p=0.003$ ;  $n=133$ ) and those with bilateral lesions (9.83;  $p=0.028$ ;  $n=33$ ) had significantly greater reflected FSD values than in people with no lesion (median 8.71;  $n=329$ ) and bilateral lesions were in fact associated with the highest median reflected FSD.

### 7.4.4 PERIORBITAL, PERINASAL AND LOWER FACE REGIONS

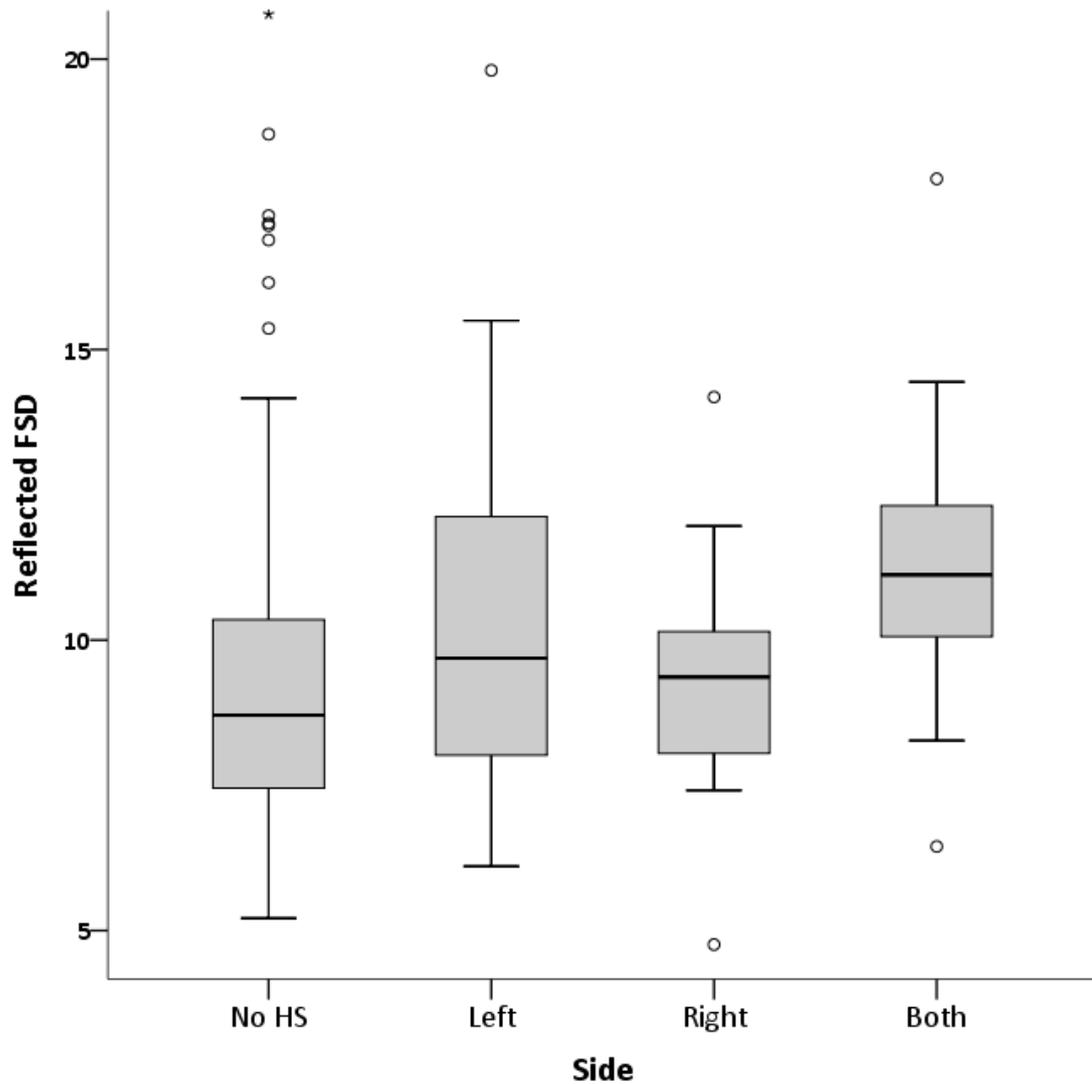
Next I assessed if particular areas of the face were significantly more asymmetric in those with developmental lesions. The periorbital region alone also showed significantly greater asymmetry using reflected FSD in those with lesions than those without (14.72 vs 13.46;  $p=0.001$ ). This was also true for the lower face alone (8.90 vs 8.32;  $p=0.006$ ) but not for the perinasal region (8.12 vs 7.66;  $p=0.092$ ).

Both the periorbital and lower face regions had a significantly greater reflected FSD in the presence of unilateral brain lesions compared to no lesion (14.72 vs 13.46;  $p=0.005$ ; periorbital, 8.8 vs 8.32;  $p=0.011$ ; lower face). In addition, the periorbital region reflected FSD was also significantly greater in the presence of bilateral brain lesions (13.85 vs 13.28;  $p=0.027$ ).

### 7.4.5 HIPPOCAMPAL SCLEROSIS AND REFLECTED FSD

I categorised every participant with HS on MRI into left HS, right HS and bilateral HS, to see if any particular subgroups were responsible for the increased face asymmetry (Figure 68). Median reflected FSD was significantly greater ( $p=0.003$ ; Kruskal-Wallis test) in people with bilateral HS (11.12;  $n=11$ ) than in left HS (9.69;  $n=46$ ), right HS

(9.36; n=36) or no lesion (8.71; n=329). However, when assessing individual face region models, there was no significant difference ( $p=0.072$  for periorbital region;  $p=0.164$  for perinasal region;  $p=0.068$  for perioral region).



**Figure 68 Comparison of face asymmetry in people with no hippocampal sclerosis (HS), left HS, right HS or bilateral HS. Reflected FSD was greater in people with bilateral HS (11.12; n=11) than in left HS (9.69; n=46), right HS (9.36; n=36) and no HS (8.71; n=329)**

Seventeen out of 93 subjects were included in a previous study and were diagnosed as having febrile seizures and mesial temporal lobe epilepsy in addition to HS (Kasperavičiūtė et al., 2013). These subjects did not have a significantly different reflected FSD compared to others with HS (9.33 vs 9.60;  $p=0.518$ ).

#### 7.4.6 ANALYSIS AFTER EXCLUSION OF ACQUIRED FACIAL DEFORMITY

Sixty-one participants with no developmental lesion had a visible acquired image deformity, as did 21 participants with any developmental brain lesion. They were excluded in a subsequent analysis of my findings due to possible confounding of the results. Using the whole face, median reflected FSD was still significantly greater in people with any developmental lesion, people with HS and people with MCD when compared to people with no lesions (Table 36).

The median reflected FSD of all those with an acquired image deformity was 9.77. I also looked at how much a specific deformity may affect reflected FSD. There were 15 people with broken noses in their clinical history, who had also been excluded for nose scars or nose deviation on blinded review. In these 15 people, median reflected FSD was 9.29 for the whole face (vs 8.51 without deformity), 10.1 for the perinasal region (vs 7.55 without deformity), 15.1 for the periorbital region (vs 13.3 without deformity) and 9.04 for the perioral region (vs 8.26 without deformity). Calculations were also made for the periorbital, perinasal and lower face models after excluding images with acquired deformity. These are shown in Table 37.

Periorbital DSM (Eyes6)							
Lesion type	None	BH	DNT	FCD	MCD	PVH	TS
Number	315	5	9	25	9	12	5
Median reflected FSD	13.28	13.16	15.1	13.87	15.53	13.29	13.71
p value	N/A	0.936	0.351	0.093	0.02	0.849	0.587
Perinasal DSM (Nose6)							
Lesion type	None	BH	DNT	FCD	MCD	PVH	TS
Number	301	7	9	24	9	12	5
Median reflected FSD	7.55	7.33	8.39	8.35	8.1	6.64	7.92
p value	N/A	0.903	0.24	0.455	0.127	0.361	0.293
Perioral DSM (Mouth6)							
Lesion type	None	BH	DNT	FCD	MCD	PVH	TS
Number	307	7	10	21	8	12	5
Median reflected FSD	8.26	8.92	8.23	8.28	9.52	9.31	7.19
p value	N/A	0.697	0.779	0.241	0.115	0.153	0.962

**Table 37 Median reflected FSD is shown for participants with no acquired image deformity in each of the three localised face models – periorbital, perinasal and the lower face. Values for significance were on comparison with participants who had no brain lesions. The number in each group is after removing all images with visible acquired deformity. None of the models found a significantly different reflected FSD in any malformation group, except for the periorbital model in detecting people with MCD. BH = band heterotopia, DNT = dysembryoplastic neuroepithelial tumour, FCD = focal cortical dysplasia, FSD = face shape difference, HS = hippocampal sclerosis, MCD = other malformation of cortical development, PVH = periventricular heterotopia, TS = tuberous sclerosis**

#### 7.4.7 REFLECTED FSD AND ASSOCIATION WITH AGE AND SEX OF PARTICIPANTS

To investigate if the sex of a person affected asymmetry, I calculated the reflected FSD separately for women (n=185) and men (n=144) without any developmental brain lesion. Men had a significantly greater reflected FSD than women (9.25 vs 8.33;  $p < 0.001$ ).

Therefore, reflected FSD was then calculated separately by sex for those with any developmental brain lesion, including hippocampal sclerosis. In women only, reflected FSD was still significantly greater in those with lesions than those without (9.29 vs 8.32;  $p = 0.004$ ; n=277). In men only, reflected FSD was no longer significantly greater (9.75 vs 9.25;  $p = 0.066$ ; n=218).

Age and reflected FSD were compared in people without lesions. There was a weak positive correlation between the two factors ( $\rho=0.253$ ;  $n=329$ ). Using simple linear regression, the line of best fit suggested that each year of age led to an increase in reflected FSD of 0.046. As described earlier, participants with developmental brain lesions had a greater median age than those with no lesion (Table 36). Therefore, I standardised reflected FSDs to an arbitrary age of 30.0 years. This was done by using the linear regression formula above to adjust each individual's reflected FSD for the age as number of years older or younger than 30.0. A revised analysis showed that the reflected FSD in those with developmental lesions was still greater than in those with no lesion (9.14 vs 8.56;  $p=0.007$ ;  $n=495$ ).

#### 7.4.8 REFLECTED FSD AND PRESENCE OF STRUCTURAL VARIANTS

It was found earlier that people with structural variants have an atypical face shape using dense surface modelling in Chapter 4. Here it was assessed if such people also had increased face asymmetry. There were 224 people who had not been found to have structural variants and 21 people who had structural variants. I found no significant difference in reflected FSD between people with structural variants and those without (10.17 vs 9.38;  $p=0.127$ ;  $n=245$ ).

#### 7.4.9 REFLECTED FSD AND BODY MASS INDEX

Body mass index may affect a participant's facial shape, especially in epilepsy due to the effects of medications, altered lifestyle and underlying syndromes. I repeated the analysis in only those subjects with a body mass index in the normal range (greater than or equal to 18.5 kg/m<sup>2</sup> and less than 25 kg/m<sup>2</sup>; Flegal et al., 2013). A total of 145 people were included and those with developmental brain lesions ( $n=43$ ) had a greater median reflected FSD than those without (9.19 vs 7.97;  $p=0.007$ ).

The difference was also significant for perioral lower face reflected FSD (8.80 vs 7.63;  $p=0.004$ ), but not for periorbital or perinasal reflected FSD.

#### 7.4.10 HANDEDNESS AND REFLECTED FSD

Handedness was recorded for 428 out of 495 subjects. Six people stated they were ambidextrous and were excluded. In the remaining 422 subjects, there were 285 with no malformation of cortical development, of whom 234 were right-handed and had a significantly smaller reflected FSD than those who were left-handed (8.46 vs 9.18; Mann-Whitney U test;  $p=0.032$ ; Table 38).

In 137 people with a developmental brain lesion, 106 had a unilateral malformation and 31 had bilateral malformations. There were 22 left-handed people with the remainder being right-handed. Right-handed people with any unilateral malformation had a significantly smaller reflected FSD than left-handed people (9.47 vs 10.7; Mann-Whitney U test;  $n=106$ ;  $p=0.016$ ; significant after Bonferroni correction). No significant difference was seen in reflected FSD between people who had a unilateral malformation ipsilateral to their dominant hand compared to those who had a unilateral malformation contralateral to their dominant hand (9.59 vs 9.69; Mann-Whitney U test;  $n=106$ ;  $p=0.714$ ). Data are also shown for the largest single group of malformations, which was people with hippocampal sclerosis. It appears that FSD is also lower in this group in right-handed people. Data were not assessed for any other single group of malformations given the small numbers involved; for example, there were only 1-3 left-handed people for any other individual malformation.

	Handedness	
	Right	Left
<b>Location of any brain lesion</b>		
No lesion	8.46 (234)	9.18 (51)
Right hemisphere	9.47 (45)	10.50 (8)
Left hemisphere	9.35 (46)	11.11 (8)
Bilateral	9.27 (24)	10.17 (6)
<b>Hippocampal sclerosis</b>		
Right hemisphere	8.93 (23)	9.76 (5)
Left hemisphere	9.43 (30)	12.16 (6)
Bilateral	10.45 (6)	12.89 (3)

**Table 38** *The relationship of handedness of subjects to reflected FSD in the presence of developmental brain lesions and more specifically for hippocampal sclerosis. In those with no lesions on MRI, right-handed subjects had a significantly smaller reflected FSD than in left-handed people (8.46 vs 9.18; p=0.032).*

## 7.5 DISCUSSION

This study uses an established method of quantifying face shape variation, which has been used in people with epilepsy and other disorders. The method has been used to assess facial asymmetry specifically, as judged by the measurement of reflected FSD between a face image and its mirror image. The same measurement was also previously used in a study of children with autism spectrum disorders (Hammond et al., 2008). In the current cohort, I found that developmental brain lesions, as a collective group, may be associated with an increase in facial asymmetry. In particular, hippocampal sclerosis and unclassified malformations of cortical development were most easily detected. The presence of bilateral brain lesions appeared to be associated with greater facial asymmetry. By contrast, putatively pathogenic structural variants were not associated with facial asymmetry. There is also a significant effect of age and sex on facial asymmetry. Finally, developmental brain lesions were still associated with increased facial asymmetry after attempting to control for a number of factors, including age, sex, body mass index and acquired facial deformity.

Using the whole face DSM was found to discriminate asymmetry best, followed by the periorbital DSM and then by the other two DSMs. This order is in keeping with the

results found in people with pathogenic structural variants earlier. The finding could suggest that the nose, maxilla and lower jaw show less variation in symmetry for anatomical reasons, such as the need for adequate dental occlusion to enable chewing. However, it could also suggest that these structures are less influenced by the forebrain during development, and hence there is less asymmetry of these in people with developmental brain lesions. In terms of the mechanism by which developmental abnormalities in the forebrain may increase facial asymmetry, there may be disturbed and asymmetrical molecular signalling. It is unlikely that a direct physical effect of the malformation affects the face, given that many lesions, such as hippocampal sclerosis, do not affect overall brain size and shape significantly.

Facial asymmetry has been described as part of the wider phenomenon of fluctuating asymmetry, which refers to small deviations in symmetry in an organism and has been thought to be correlated to the degree of “developmental instability” (Dongen, 2006). Developmental instability is a concept describing the sensitivity of an organism to pathogens, toxins or genetic mutations. In humans, the known association of facial attractiveness and symmetry has been justified as an evolutionary means of assessing developmental instability in potential mates (Rhodes, 2006). Since the 1970s, facial asymmetry in particular has been associated with chromosomal disorders, such as Down syndrome and trisomy 14 (Thornhill and Møller, 1997).

With stereophotogrammetry and the increased sophistication of facial asymmetry measurement, other studies have been conducted. One study of 60 Chinese adults used landmarks on 3D face images and the deviation of these from facial planes in three dimensions (Huang et al., 2013). In their study, nasion was defined as being in the vertical midline of the face and a triangle connecting nasion and both exocanthions was defined to be in the horizontal plane. These definitions mean that asymmetry affecting nasion or the exocanthions are not adequately detected and also that such asymmetry

could lead to errors in the planes used, leading to errors in calculations of their facial asymmetry index. Even without this error, this landmark-based method is incapable of detecting facial asymmetry that does not affect landmarks.

The largest stereophotogrammetry study has used data from the Avon Longitudinal Study of Parents and Children (Pound et al., 2014). The authors of the study found that in children, increased facial asymmetry was not associated with increased periods of ill health, defined by the presence of symptoms such as earache or vomiting, but neurodevelopmental disorders were not specifically assessed. They used a landmark-based geometric morphometric approach. One of their secondary findings was that IQ, which had been measured in some children, was significantly correlated with facial symmetry, although it appears to be very weak with a Pearson's correlation coefficient of -0.056 (Pound et al., 2014). The measurement used by this group is a correlation coefficient and not an absolute value, and also it had to be controlled for face size. Both of these suggest that reflected FSD may be a better marker of face asymmetry, although their study was in children not adults.

On the other hand, another study using stereophotogrammetry and dense surface modelling found that facial asymmetry was increased in boys with autism spectrum disorder as well as unaffected mothers of such boys (Hammond et al., 2008). The findings in the current study lend support to the theory that abnormal development of the brain, using the phenotype of autism or cortical malformations, leads to increased facial asymmetry. Bilateral malformations may be an indirect marker of the extent of abnormal brain development, which in turn could be associated with the extent of facial asymmetry.

Facial asymmetry also existed in people with epilepsy without developmental brain lesions. It is likely that even in people without epilepsy or any other neuropsychiatric disorder, there will be some asymmetry. Farkas and colleagues found asymmetry of up

to 3mm for landmarks in one of the earliest studies in North American people (Farkas and Cheung, 1981). In one of the two previous stereophotogrammetry studies, such asymmetry existed in control boys (Hammond et al., 2008) and in the other study, the extent of asymmetry was not described; only correlations are stated (Pound et al., 2014). A study using laser scanning found an “average” difference of 3mm between facial measurements on the left and right side in one study of healthy Finnish adolescents (Djordjevic et al., 2013). This normal asymmetry could theoretically be lost in certain developmental syndromes, although there was no trend or suggestion of decreased reflected FSD in those with any particular cortical malformation.

Facial injury and trauma is potentially a significant confounding factor in this study. Unlike other factors such as age and BMI, facial injury/trauma is more likely to cause unilateral differences to the face shape and hence probably increase facial asymmetry. The two previous studies of face asymmetry using stereophotogrammetry do not mention or exclude subjects with acquired face deformity (Hammond et al., 2008; Pound et al., 2014). In this chapter, I first excluded people with visible acquired image deformity on blinded review. This led to two findings. Firstly, reflected FSD is still able to distinguish people with mixed or undefined malformations of cortical development and of hippocampal sclerosis alone. Secondly, those people with visible acquired facial deformity do have a corresponding increase in reflected FSD of approximately 1 unit from 8.71 to 9.77. Clinical history was also available but was not used to avoid subjective bias in excluding people with injuries that may not have affected face shape. However, the clinical history and the blinded review were both used in conjunction to identify 14 people with no malformation of cortical development and with a history of nose fracture that was visible on the face image. In these 14 people, median reflected FSD was also greater, especially in the perinasal and periorbital regions as expected. These results mean that face shape analysis for symmetry should take into account face injury and trauma in future.

Another finding was that right-handed people had a smaller reflected FSD in all groups analysed compared to left-handed people. There was no observed difference between lesions ipsilateral to or contralateral to the dominant hand of subjects. It is not clear why left-handedness is associated with increased facial asymmetry in my study. A higher prevalence of left-handedness has been documented in people in autism spectrum disorders, schizophrenia and Down syndrome (Lindell and Hudry, 2013). A small study of children with both focal cortical dysplasia and hippocampal sclerosis were more likely to be left-handed than in the general population (Maulisova et al., 2016). It may be that left-handed people had undetected malformations of cortical development, which led to increased facial asymmetry. Another possibility is that left-handedness and malformations of cortical development are both independent predictors of increased facial asymmetry.

There are a number of limitations of this study. The most important limitation is that the study had very small numbers of cases for individual brain lesions other than HS. The strongest association of any type of lesion with facial asymmetry was for HS, which was also easily the largest group. It may be that with larger groups of people with DNT, FCD and other types of lesions, one could assess which of these lesions may increase facial asymmetry.

I did not have full clinical phenotypes for all participants. For example, some of the subjects may have autism spectrum disorders, which are now known to be associated with facial asymmetry in boys (Hammond et al., 2008). The phenotypes may also have evolved since my data collection. Malformations of cortical development have an evolving classification increasing in complexity, such as with new genetic factors or the use of 7-Tesla MRI scans (Guerrini and Dobyns, 2014). This also means that lesions detected on MRI may be misclassified. Only 22 of the 166 people with developmental brain lesions had undergone surgery and therefore there was no pathological

confirmation of diagnosis in the majority of cases. Detection requires neuroradiological expertise as well as appropriate high resolution sequences. Indeed, in one of our centres, the use of a higher resolution MRI scanner enabled an increased diagnostic yield for DNT, FCD and HS (Winston et al., 2013). Another limitation is that despite the largely objective method used, the landmarking of face images and the assessment of acquired image deformity remain subjective steps in which the operator must judge the best placement. Blinding and high reproducibility may not be sufficient to eliminate bias or error at this step.

Our findings with developmental brain lesions are in a group of people with epilepsy and it is not clear if the results could be generalised to people with other developmental brain lesions in the absence of epilepsy. This caution over generalisability is also true of my findings that face asymmetry may be increased in men and with age.

## 7.6 CONCLUSION

This study supports the hypothesis that abnormal development of the brain may result in abnormal development of the face. There is greater asymmetry of face shape in people with epilepsy with developmental brain lesions than in people with epilepsy without a developmental brain lesion. The presence of bilateral brain lesions is associated with greater asymmetry than unilateral lesions. Hippocampal sclerosis alone is associated with increased face asymmetry. Right-handed subjects appear to have less face asymmetry than left-handed subjects. Facial injury needs to be carefully assessed as this increases face asymmetry significantly. Further research is needed to elucidate the molecular mechanisms whereby this occurs in lesions such as HS.

## CHAPTER 8: FACE SHAPE IN *HNF1B* HAPLOINSUFFICIENCY

### 8.1 INTRODUCTION

The hepatocyte nuclear factor 1-beta (*HNF1B*) gene is a transcription factor that is important for organogenesis. Mouse studies have shown that *HNF1B* deficiency is lethal and the gene is expressed during embryonic development of the visceral endoderm (Coffinier et al., 1999), including the gut (D'Angelo et al., 2010), kidneys (Hiesberger et al., 2004), pancreas (Haumaitre et al., 2005), biliary system and liver (Coffinier et al., 2002). In humans, heterozygous mutations or deletions of the *HNF1B* gene can lead to renal cysts or malformations (Edghill et al., 2006), pancreatic atrophy with exocrine and endocrine insufficiency (Bellanné-Chantelot et al., 2004; Horikawa et al., 1997), hepatobiliary dysfunction (Roelandt et al., 2012), and genital tract malformations (Oram et al., 2010).

More recently, deletions involving the *HNF1B* gene have also been associated with autism, schizophrenia, developmental delay and facial dysmorphism (Moreno-De-Luca et al., 2010; Nagamani et al., 2010). Seizures and structural brain abnormalities have been associated with both deletions and duplications of 17q12, which encompasses the *HNF1B* gene (Mefford et al., 2007; Nagamani et al., 2010). Such structural brain abnormalities include frontal gliosis, hippocampal hyperintensities, and hippocampal atrophy (Moreno-De-Luca et al., 2010; Nagamani et al., 2010). In one study, those with deletions were thought to have consistent facial features, which included arched eyebrows, slanting palpebral fissures, and a depressed nasal bridge (Moreno-De-Luca et al., 2010). However, in others with deletions, there was no facial dysmorphism or there were only inconsistent features in a minority, such as a triangular face, epicanthal folds, low-set ears, or micrognathia (Edghill et al., 2008; Moreno-De-Luca et al., 2010; Nagamani et al., 2010). No obvious facial dysmorphism or structural brain abnormality has been reported in people with *HNF1B* point mutations (Bellanné-Chantelot et al.,

2004; Edghill et al., 2006). Our group has described one person with epilepsy and a microdeletion encompassing the *HNF1B* gene (Kasperavičiūtė et al., 2011), but a review suggests that epilepsy only occurs in isolated cases and may not be linked to this gene specifically (Bockenbauer and Jaureguiberry, 2016).

It is possible that *HNF1B*, a gene that directs embryonic development and is expressed in the developing mammalian brain (Makki and Capecchi, 2011), may affect facial morphology. As explained in Chapter 1, facial development is known to be dependent on signals from the developing forebrain and aberrations in one may be associated with aberrations in the other (Marcucio et al., 2015). Subtle facial dysmorphism or asymmetry may not have been considered in earlier case series of *HNF1B* mutations.

Three-dimensional stereophotogrammetry and dense surface modelling have been previously used to analyse facial morphology and symmetry in genetic and neuropsychiatric syndromes (Hammond et al., 2004, 2005; Cox-Brinkman et al., 2007; P. Hammond et al., 2008). One finding was of significantly increased facial asymmetry in autism spectrum disorders, which may be due to unidentified genetic factors (P. Hammond et al., 2008). I have also applied dense surface modelling to explore consequences of underlying SVs (including deletions), syndromes and brain lesions in people with epilepsy in chapters 5, 6 and 7, as well as in collaboration in other studies (Coppola et al., 2013; Novy et al., 2012).

Here I used dense surface models (DSMs) to further investigate facial shape in a pilot study of children and adults with mutations or deletions of the *HNF1B* gene.

## 8.2 HYPOTHESIS

I hypothesise that as a gene implicated in neurodevelopmental disorders and expressed in the developing brain, *HNF1B*, is important in facial development and that FSD is able

to detect a significantly atypical or asymmetric face shape in people with mutations or deletions in this gene.

## 8.3 METHODS

### 8.3.1 RECRUITMENT, IMAGE CAPTURE AND FACE SHAPE ANALYSIS

Subjects and their relatives were recruited during a support group event at the UCL Institute for Child Health as detailed below and in Chapter 2 (group 7). Inclusion criteria were all subjects with a mutation or deletion of the *HNF1B* gene. Exclusion criteria were those people who could not be compared to reference subjects due to age (>52 years for men; >54 years for women) or ethnicity (only white Europeans included).

This is the only study in the thesis that compared the group of interest to the group of reference subjects, instead of other people with epilepsy. It was not appropriate to compare this group to any of the subjects with epilepsy that were recruited during the course of this research because the presence of epilepsy may be a confounder for multiple reasons outlined in Chapter 1. Not all of the subjects here with *HNF1B* mutations or deletions had epilepsy. Face images of reference subjects were therefore used for comparison, and these subjects were recruited and landmarked by Peter Hammond and Mike Suttie as described in Chapter 2. I have shown that there is good inter-operator reproducibility in Chapter 3. The reference subjects had no known syndrome but did not undergo genotyping. DSMs were created as described in Chapter 2. The DSMs included reflected face images used to calculate reflected FSD as well as FSD.

### 8.3.2 GENETIC ANALYSIS

One of three methods had been used to detect *HNF1B* gene abnormalities. Two of these methods were all applied by and analysed by a collaborating group at the University of Exeter comprising Dr Coralie Bingham, Dr Rhian Clissold and Prof Sian Ellard. Gene

sequencing was performed as previously described (Oram et al., 2010). In summary, nine exons and the minimal promoter were amplified using polymerase chain reaction, with tagging of sequence-specific primers, and then were sequenced in forward and reverse directions using the Big Dye Terminator Cyclor Sequencing Kit (Applied Biosystems, Warrington, UK). Sequences were analysed using Mutation Surveyor Software (SoftGenetics, LLC, State College, PA) and checked against published polymorphisms and mutations.

Multiplex ligation-dependent probe amplification (MLPA) dosage analysis was performed in subjects who had no mutations on gene sequencing, as described earlier (Edghill et al., 2008). In summary, nine exon-specific probes were used with the MEN1 MLPA Kit (#P017; MRC Holland, The Netherlands) in a fluorescently labelled multiplex ligation/PCR reaction. The data were analysed using GeneScan and Genotyper version 2 analysis software (Applied Biosystems, Warrington, UK).

The laboratories were the North East Thames Regional Genetics Service and the Molecular Genetics Laboratory, Royal Devon and Exeter NHS Foundation Trust, Exeter.

Oligonucleotide array comparative genomic hybridisation (aCGH) was performed in one case, using the Nimblegen 135K microarray (Roche Nimblegen) and analysed using CGH Fusion software (Infoquant, UK). This case was analysed at the clinical genetics laboratory in Great Ormond Street Hospital for Children NHS Trust.

All testing was performed in accordance with manufacturers' instructions.

### 8.3.3 STATISTICAL ANALYSIS

The Mann-Whitney U test was used to compare people with a defect of the *HNF1B* gene to the control population. Analysis was conducted using SPSS (versions 19-23, SPSS Inc; Chicago, IL, USA).

## 8.4 RESULTS

### 8.4.1 SUBJECT POPULATION

Eight probands with *HNF1B* gene abnormalities were recruited, along with two unaffected relatives (Table 39). One proband has already been investigated for facial changes and is part of earlier analyses (see Chapter 2) and is described in the literature (Kasperavičiūtė et al., 2011). Gene sequencing identified point mutations in four people, including one frameshift mutation due to loss of a single base, and MLPA or aCGH found deletions in the remaining four. Their images were compared in four DSMs each with 197 control images, comprising 98 male subjects and 99 female subjects (Table 40).

	Subject	Relation	Age (yrs)	Sex	HNF1B genotype	Method used	Matched face shape difference	Reflected face shape difference
01	Proband	Brother of 02	11.3	Male	HNF1B exon 9 del	MLPA	7.31	6.76
02	Proband	Brother of 01	14.2	Male	HNF1B exon 9 del	MLPA	10.09	10.28
03	Proband	-	21.4	Male	c.511T>C, p.W171R (missense mutation)	Sequencing	9.87	10.31
04	Proband	-	29.2	Male	17q12 del	aCGH	12.11	9.44
05	Proband	-	35.5	Male	c.826C>T, p.R276* (nonsense mutation)	Sequencing	9.47	11.84
06	Proband	-	35.8	Female	Met1_Trp557del	MLPA	10.87	12.85
07	Proband	-	43.5	Male	c.832delG, p.V278fsdelG (frameshift mutation)	Sequencing	11.48	10.44
08	Proband	-	46.9	Female	c.451T>C, p.S151P (missense mutation)	Sequencing	14.23	21.81

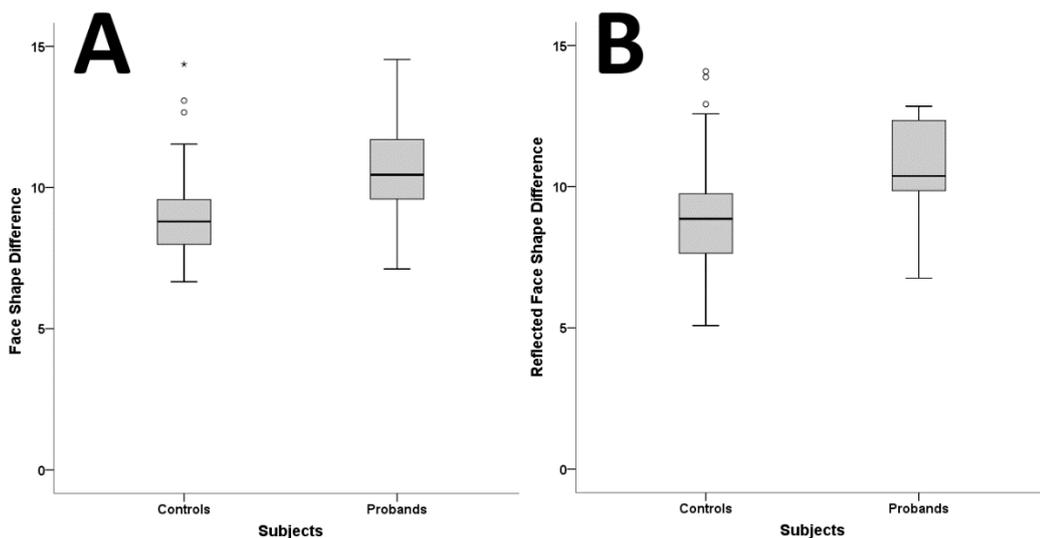
**Table 39** Details for all subjects with *HNF1B* abnormalities, including age, sex, type of *HNF1B* gene defect and the results for face shape difference.

	Control subjects	Subjects with <i>HNF1B</i> abnormalities
Number	197	8
Age in years (median, IQR)	28.6 (20.7)	32.4 (18.1)
Sex (male, %)	98 (50%)	6 (75%)
Matched FSD (median)	8.79	10.45
Significance	N/A	p=0.007
Reflected FSD (median)	10.38	8.86
Significance	N/A	p=0.007

**Table 40 Characteristics of the control group and the group of subjects with *HNF1B* gene abnormalities. Both FSD and reflected FSD were significantly increased in subjects with *HNF1B* abnormalities.**

#### 8.4.2 FSD AND REFLECTED FSD IN PEOPLE WITH *HNF1B* GENE ABNORMALITIES

The median FSD for the eight probands was significantly higher than in the 197 controls (Figure 69 panel A; 10.45 vs 8.79; p=0.007). Three control subjects were outliers and excluding them did not affect the significance of the results. Reflected FSD was then calculated from the set of both original and reflected face surfaces. Median reflected FSD was also significantly higher in the eight probands than in controls (Figure 69 panel B; 10.38 vs 8.86; p=0.007).



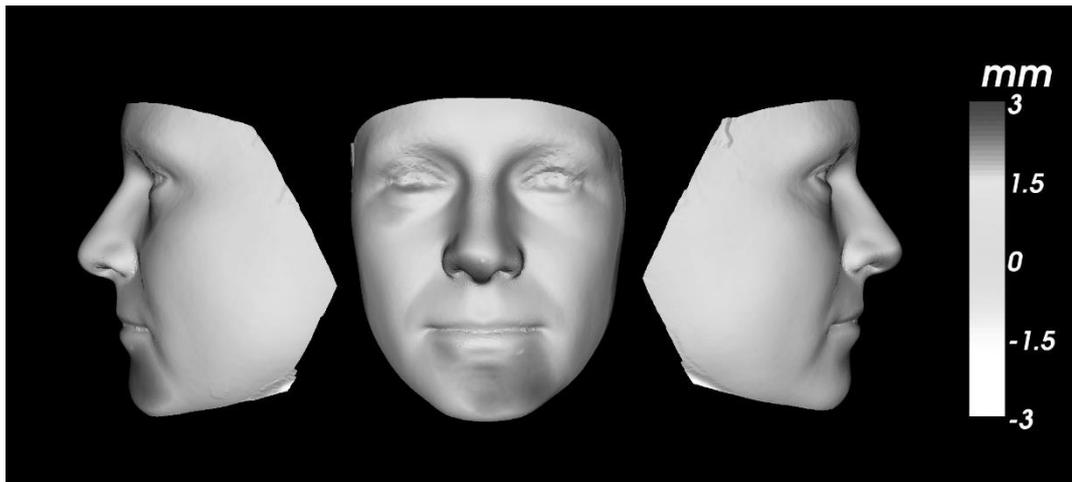
**Figure 69 (A) Box plots of the median, interquartile range and range show FSD is significantly greater in people with *HNF1B* gene abnormalities compared to controls (median 10.45 vs 8.79; p=0.007). (B) Similar box plots for reflected FSD also show it is significantly greater in people with *HNF1B* gene abnormalities (median 10.38 vs 8.86;**

***p=0.007). Outliers over 1.5 or 3 times the interquartile range from the upper quartile are shown in circles or asterisks respectively. Excluding all outliers maintains significance.***

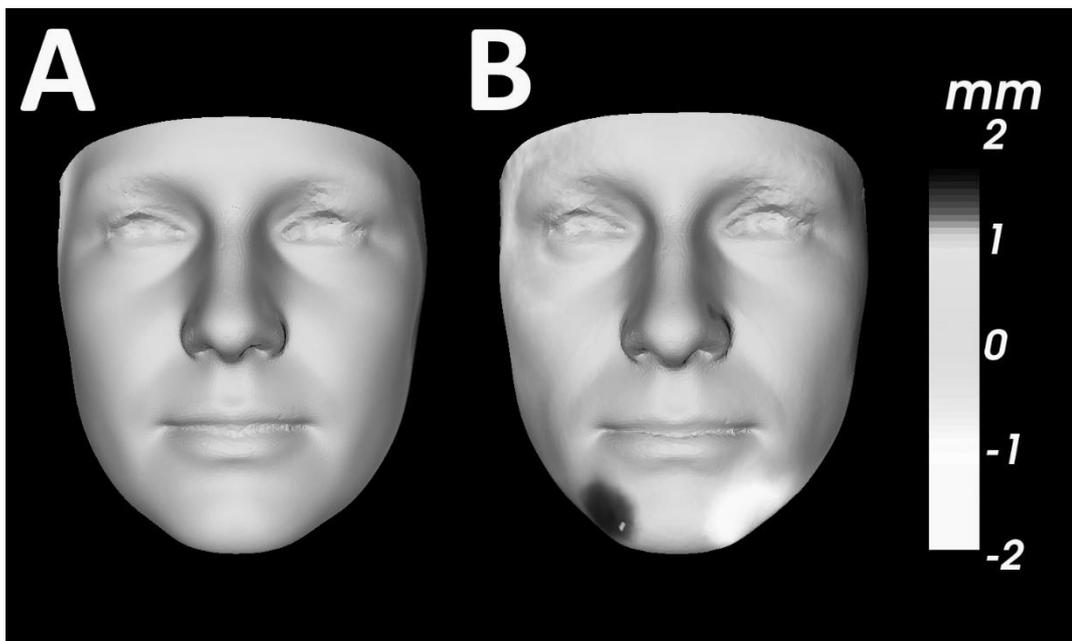
There was significant positive correlation between FSD and reflected FSD in reference subjects (n=197,  $\rho=0.65$ ;  $p<0.001$ ) but not in probands (n=8,  $\rho=0.43$ ;  $p=0.29$ ). In addition, there was no significant difference in FSD and reflected FSD between probands with gene deletions (n=4) and those with point mutations (n=4).

### 8.4.3 VISUAL INSPECTION OF MEAN FACES

The mean face surface of the eight probands was synthesised using the mean for every principal component, and compared with the mean face surface of the control subjects (Figure 71). This did not show any clear changes in the overall face surface shape of probands, but there was outward displacement of 1-2mm at the lower eyelid, the nasal tip, and the mandible. Note that this visual group comparison, unlike FSD and reflected FSD, does not standardise for age and sex differences between controls and probands. To investigate the nature of facial asymmetry, the mean face surface was compared to the mean reflected face surface, separately for controls and for probands. In the graphical comparisons, probands appear to show more asymmetry over the anterolateral surface of the mandible and the zygoma (Figure 71).



**Figure 70** The mean face of all eight probands is shown using a white-grey-dark grey heat map, representing inward-null-outward displacement in millimetres along the surface normal, compared to the mean face of all control subjects. This was viewed on a computer monitor with red-green-blue replacing white-grey-dark grey to identify changes more easily. There is no marked difference seen in the mean face. However, there is a suggestion of increased displacement from the average at the palpebral fissures, the nasal tip and the mandible. Note this only looks at the mean faces in probands and may not represent individual face shape changes from control face images.



**Figure 71** The mean control face surface is shown on the left (A) and compared to its reflected face surface using grayscale heat maps. The equivalent comparison for the mean face surface of all eight probands and its reflection is shown on the right (B). The heat maps are shaded white-grey-black for inward-null-outward displacement along the surface normal respectively. In probands, there appears to be noticeable facial asymmetry over the anterolateral surfaces of the body of the mandible. No discernible asymmetry is seen in control faces.

## 8.5 DISCUSSION

Dense surface modelling showed that in people with *HNF1B* deletions or mutations, there was a significantly more atypical face shape with increased asymmetry, when compared with matched controls. However there appeared to be no common facial features underlying this difference in shape and symmetry.

Previous studies of *HNF1B* mutations found no common facial features or dysmorphism in those affected. Two previous studies provided detailed phenotypes in people with *HNF1B* deletions and found that fourteen people had no facial dysmorphism and one had prognathism (Edghill et al., 2008; Nagamani et al., 2010). One study of eighteen people with *HNF1B* deletions described facial phenotypes of nine people (Moreno-De-Luca et al., 2010). All nine had a 1.4Mb deletion of 17q12, which contained 15 genes, including *HNF1B*. Of these, seven were thought to have cranial shape abnormalities, which included a high forehead, scaphocephaly, and frontal bossing. Four had ear abnormalities, such as fleshy earlobes or prominent lobes. Eight had arched eyebrows, and six had upslanting or downslanting palpebral fissures – “mildly” or “slightly” in four of these cases. Nasal abnormalities included a depressed nasal bridge in seven cases and a tubular shape in four cases. The mouth was thought to be dysmorphic in four cases. The authors concluded that there were similar mild facial dysmorphic features in all nine cases. Such dysmorphism was not noted in the subjects here, and may reflect the involvement of other genes on facial shape, or the small sample size.

Supporting the involvement of other genes, there was one subject, 04, in this study with a 17q12 deletion containing genes other than *HNF1B*. He had a 2.1Mb deletion with 16 genes in the deletion interval, and his case, including dysmorphic features, has been reported previously (Kasperavičiūtė et al., 2011). It was hypothesised that haploinsufficiency of another 17q12 gene, *ACACA*, contributed to mild facial

dysmorphism. He was also noted to have overlapping facial features with the subjects in the above study by Moreno-De-Luca and colleagues, who also have a heterozygous deletion of the *ACACA* gene (Moreno-De-Luca et al., 2010). In the current study, FSD for subject 04 was higher than the median for the probands, although this was not true for reflected FSD.

In Chapter 4, FSD detected a more atypical face shape in people with epilepsy and genomic changes, but did not find common features in those affected. In contrast, the method has identified common facial features in discrete syndromes or genetic disorders, including those with no obvious detectable dysmorphism (Cox-Brinkman et al., 2007; Hammond et al., 2005). Haploinsufficiency of *HNF1B* is one genetic disorder but in this small study, I found no similarity in facial shape amongst people with the condition. One hypothesis is that *HNF1B* may closely regulate the development of the face when expressed as a transcription factor in the embryo, such that in its absence there is greater variation in the resulting face shape, including increased asymmetry.

There was no difference in facial shape and symmetry between the four probands with a deletion compared to the four with point mutations. This agrees with studies that have found similar clinical phenotypes in people with deletions and mutations (Edghill et al., 2008).

Increased facial asymmetry in the probands may be relevant to the known association of *HNF1B* with autism spectrum disorders. Children with autism spectrum disorders were earlier found to have increased facial asymmetry using dense surface modelling and this was partly due to right periorbital prominence (Hammond et al., 2008). These children were not genotyped and some may have had *HNF1B* mutations or deletions; in this study, subjects were not formally evaluated for features of autism spectrum disorder.

A limitation of the findings is that this was a small study because of difficulty in recruitment of affected subjects. Two patients were also brothers and thus shared facial features in them could have influenced the mean face surface of all subjects with *HNF1B* gene abnormalities. Dense surface modelling in a larger group of affected individuals may help to show common features and also any differences between those with mutations and those with deletions. The reference control population used was not genotyped and was also recruited separately by different operators and sometimes a different camera. These may all have introduced errors in calculating FSD as explored in my earlier reproducibility studies.

## 8.6 CONCLUSION

In conclusion, the findings here support a role of *HNF1B* in regulation of facial morphogenesis. Whilst it would seem that there are no shared facial features amongst people with deleterious changes in *HNF1B*, the sample size is small, and further cases will need to be evaluated for definitive conclusions.

## CHAPTER 9: DISCUSSION

To my knowledge, this is the first time that stereophotogrammetry and dense surface modelling have been used to study people with epilepsy. In addition, it is the first study to use a single metric, Face Shape Difference (FSD), to characterise if a face shape is atypical.

### 9.1 STEREOPHOTOGRAMMETRY

I have shown here that stereophotogrammetry can be successfully applied to a population of people with epilepsy. In total, 869 people were recruited for my study over two years. Subjects included adults and children. People with severe intellectual disability were recruited as were subjects who were wheelchair-bound. The portability of the camera allowed recruitment at different sites, including hospitals in Europe and a support group event in London. Resources required are minimal. The camera has been used in a space of approximately six square metres and required a tripod, chair, table and a power source.

The Canfield Vectra CR10 camera system used here is a commercial stereophotogrammetry system devised for clinical use, which is comparable to other such camera systems (Tzou et al., 2014). This camera has been previously used in studies of face and breast shape (Cox-Brinkman et al., 2007; de Menezes et al., 2010; Ghoddousi et al., 2007; P. Hammond et al., 2008; Metzler et al., 2014; Othman et al., 2013; Quan et al., 2011; Roostaeian and Adams, 2014; Rosati et al., 2014, 2012, 2010; Rossetti et al., 2013). Two reproducibility studies looked at error in distances and angles between facial landmarks (de Menezes et al., 2010; Ghoddousi et al., 2007).

In this thesis, my camera was used in a different environment, such as different lighting conditions. It was used in a different population of subjects, including those with intellectual disability and children. There are differences in image capture due to

seating position, instructions given to subjects and how images were judged to be acceptable. There are then also differences in landmarks used. In addition, the actual camera is a complex machine that may have had technical faults. For all of these reasons, I initially conducted reproducibility studies.

## 9.2 REPRODUCIBILITY STUDIES

Inter-operator and intra-operator reproducibility error was assessed for landmark positioning by using calculated distances between landmarks, with the results being comparable to previous studies. Individual landmarks varied in terms of reproducibility error and the least reproducible landmarks were similar to those found previously too. As far as I am aware, for the first time in stereophotogrammetry studies of the human face, I then also assessed the reproducibility error due to head positioning relative to the camera and camera calibration, a step that is needed in every imaging session to assemble a 3D image. Calibration did not affect reproducibility. However, I showed that head position, especially lateral rotation of the neck (i.e. turning to look left or look right) beyond 5 degrees from the midline may adversely affect image capture. This was controlled for by asking subjects to fixate on an object vertically above the centre of the camera and checking images for fixation. Such visual inspection alone appeared to be adequate in terms of reproducibility of landmarks and FSD. However, it is a finding for other stereophotogrammetry studies to note.

Face expression was also assessed for four expressions: mouth open, lip puckering, smiling and neutral expression. Other studies have only looked at reproducibility of landmarks in subjects with these expressions. I sought to see how much these expressions affected landmark positions and FSD. It was shown that FSD is sensitive to facial expressions, even those which may not appear to affect FSD. For example, the periorbital region FSD is significantly increased by subjects smiling, even though landmarks around the eyes do not seem to change significantly.

Therefore, there are three aspects to consider in future studies: landmarking, head positioning and face expression. Landmarking is the predominant subjective step in the whole process of stereophotogrammetry and dense surface modelling. When the same images are landmarked a second time and included in a different DSM, correlation of FSD between the original and repeat images is imperfect:  $\rho=0.90-0.94$ . For face surfaces to be used in databases for comparison, landmarking and subsequent FSD need to show as little variation as possible. Within the same DSM, repeat landmarking did not alter FSD greatly: the median difference in one DSM was 0.1 units, which was 0.4% of the range of FSD values seen in the DSM. Nevertheless, landmark reproducibility is likely to suffer with different operators and more than one operator is necessary for practical use of this technique in clinical settings.

Recently, algorithms have been developed to automate landmark placement based on changes in contour between surface points and the authors were able to use the algorithms for placement of 20 of the 22 landmarks used in my study (Katina et al., 2016). These 20 landmarks only omitted palpebrale superius and palpebrale inferius, and included the least reproducible landmarks: alari nasi, cheilion, gnathion, nasion and otobasion inferius. They found that automated placement showed the greatest benefit for the landmarks that were least reproducible for observers to place. They also note that their definitions for landmarks may differ from traditional definitions, as do some of my definitions (listed in Chapter 2). This does not matter for dense surface modelling providing there is consistency in landmarking. Landmarks themselves are not compared, unlike in traditional direct or indirect anthropometry.

Another option may be to avoid landmarking altogether and this is discussed in Chapter 3. The purpose of landmarking in dense surface modelling is to allow face surfaces to be registered to a base mesh, which then allows a dense correspondence of points to be created. In brain imaging, registration of brain shape may be carried out

using automated volume-based image registration, in which Procrustes analysis is employed to units of brain volume (termed 'voxels'). In orthodontics, voxels have been used with CT scan data of the facial skeleton. Indeed, voxel-based registration was at least equivalent to surface landmark-based registration in one study (Almukhtar et al., 2014). A caveat here is that only three surface landmarks were used to align surfaces. Aside from this study using CT imaging and voxels, no study to my knowledge has attempted to register face surfaces without the use of any landmarks. Theoretically, polygon-based surface registration may be possible in future though.

For head position, I attempted to see if image analysis was able to determine if a subject's head position was optimal (in Chapter 4). This was done by calculating axes across a person's face. Unfortunately, my technique to determine the angle of rotation of the head in each axis only worked for one subject after analysis of multiple previous images of the same subject at different angles, in which the subject used a chin rest for accuracy. It is difficult to use more elaborate mounts used to keep the head in one position, such as those used for slit lamp examination, visual perimetry or optical coherence tomography, because they interfere with the image capture of the face surface. Using a chin rest may be the best method to help maintain head position but needs to be used carefully to ensure the rest is outside of the area of coverage of the whole face DSM.

Facial expressions are also difficult to correct for in stereophotogrammetry studies. Few studies mention facial expression and how it was assessed. This includes stereophotogrammetry studies in children or subjects with autism (Aldridge et al., 2005; Hammond et al., 2008; Hammond et al., 2005; Heike et al., 2009), although one study recommends that, in future studies, children should have their mouth closed for image capture (Heike et al., 2009). In my own study, I asked participants to keep their lips closed, not to smile and to look ahead at the fixation target. The field of forensic

science has explored the effect of face expressions on face recognition using 2D photos and a review paper in 2010 mentioned possible strategies to counter different facial expressions for 3D face images but also noted that objective methods are “urgently required” (Smeets et al., 2010). The group advocates the creation of a database or gallery of different facial expressions or suggests that only parts of the face that do not vary with face expression are used. The latter approach is possible with dense surface modelling by using regional DSMs of the periorbital, perinasal or perioral regions, as shown in my study.

### 9.3 DENSE SURFACE MODELLING AND FSD

Dense surface modelling has been used in the analysis of face images since 2004 (Hammond et al., 2004). It has been applied in the study of 11 genetic disorders, using adults and children in predominantly white European populations (Bhuiyan et al., 2006; Cox-Brinkman et al., 2007; de Souza et al., 2013; Hammond et al., 2012a, 2012b, 2005, 2004; Heulens et al., 2013; Tobin et al., 2008). The technique has also been used in autism spectrum disorders and fetal alcohol syndrome (Hammond et al., 2008; Suttie et al., 2013).

The technique involves registration of landmarked face images, which is divided into three steps within one software program, DSMExplorer. The steps are of Procrustes analysis, thin-plate spline deformation and dense correspondence creation. After this, the final step in previous analyses was principal component analysis. Principal component analysis describes a particular set of images using principal component vectors that successively explain as much of the remaining variation in face shape as possible. The principal components are specific to that set of images, or DSM.

The step of principal component analysis itself is needed to reduce the dataset to a manageable size and in the calculation of FSD. However, this step captures only 99% of the total face shape variation explained by the principal components. By definition,

there is loss of 1% of information on face shape and this is by design. The creator of the DSMExplorer program shows that for a typical DSM, increasing the accuracy of principal component analysis to 99.9% requires approximately trebling the number of PCs and reduces the average error of a point co-ordinate by 10%. In one example, using a DSM containing 193 face images, 146 PCs were required to obtain 99.9% of the variation between face shapes with a root mean square error of 0.90mm for point co-ordinates, but only 56 PCs to obtain 99% of the variation with a root mean square error of 1.03mm (Hutton, 2004). Thus, Hutton concluded that using sufficient PCs to explain 98-99% of the variation was satisfactory. The current error is less than in studies by other groups using principal component analysis to describe face shape, in which 13-25% of information is lost in the principal component analysis step (Hennessy et al., 2010, 2007, 2002; Prasad et al., 2015).

Another drawback of principal component analysis is that it is specific to a dataset. Although in all DSMs here, the first PC describes face size, for example, the exact dimension of the vector is likely to be slightly different. This means that FSD, which is calculated from the underlying PCs, cannot be directly compared between different DSMs. A consequence of this is that all face images to be compared must be processed together using the DSMExplorer software to create a DSM that contains the images. This is partly why different DSMs of the whole face had to be created for the different chapters of this thesis. Creation of a DSM therefore limits the ease of analysis. For example, to analyse a new face image of a person with a *HNF1B* mutation, a new DSM must be created using the new subject and all of the existing people with *HNF1B* mutations or deletions, as well as all controls. Future software versions could consider allowing a new face image to be added to an existing DSM. The problem of comparing different DSMs still exists however. Other alternatives to principal component analysis, such as linear discriminant analysis and independent component analysis, are also specific to a dataset. Furthermore, linear discriminant analysis, which has been used in

DSMs before (Hammond et al., 2005), requires *a priori* classification into two categories of images in a dataset.

FSD is a novel metric used here for the first time. Unlike previous studies using DSMs and PCs, I was not looking at people with one syndrome or one condition. In the case of a single clinical syndrome, a number of methods have been used to assess face shape from visual inspection to identifying particular PCs that are different in controls and affected subjects, to linear discriminant analysis or other shape classification methods (Cox-Brinkman et al., 2007; Hammond et al., 2012a, 2012b, 2005). I was analysing the extent to which SVs could alter face shape from the expected average face shape, but I acknowledged that different SVs may exert different effects on face shape. FSD quantified the extent of the difference from the average matched face. To do this, it compared the difference in every PC between a chosen face surface and the matched average face surface.

#### 9.4 REFERENCE SUBJECTS

The average face shape was chosen as the mean of thirty faces from a reference set of faces of the same sex, of the closest age to the face being analysed. On visual inspection, the average face of thirty reference faces appeared to be smooth and no individual subject's face appeared to unduly influence the average face shape. Using fewer faces, such as ten, could mean that one distinctive reference face could distort the average face. On the other hand, using more faces would require a larger sample of reference faces. I acknowledge that the variable of the number of faces could be further explored to see if more faces can increase the accuracy of FSD.

As it was, my study was limited by age and ethnicity of the reference subjects.

Reference subjects were all white Europeans. The age ranges for the mean faces were 2-52 years for male subjects and 2.5-54 years for female subjects. These restrictions meant the exclusion of approximately 20-25% of all recruited subjects for whom FSD

was calculated, which was for the studies in chapters 4, 5 and 6. Increasing the reference sample in terms of age and ethnicity is important for larger studies using FSD. Ethnicity itself was restricted to white Europeans for this work. The historical term 'Caucasian' was not used as it is too broad, for example including people in the Middle East, and it is too vague. Ethnic variations in face shape has been studied using stereophotogrammetry by one group who thought there were detectable differences between "native" dental students in Wales, Slovakia and Hungary (Kau et al., 2010). Another group found differences between a British and a Dutch population (Hopman et al., 2014). Therefore, it may be that ethnicity should be further refined into, for example, white British people to limit differences due to ethnic origin. Another limitation of the reference subjects is that very little information was known about them. Their age, sex and ethnicity was known but there was no information about previous facial injury or surgery or BMI. None had been genotyped either and it was assumed that the reference subjects had no SVs. As mentioned before, using 30 reference subjects limits the effect that one person has on the mean face shape, but a better method in future would also be to gather more data on reference subjects.

## 9.5 DENSE SURFACE MODELLING AND THE RELATIONSHIP TO BMI AND AGEING

Face shape is strongly related to BMI and weight with over 60% of variance in BMI or weight predicted by the first three PCs of one DSM (see Chapter 6). The PCs appear to have a linear relationship with BMI and this allows one to attempt to correct for BMI. I have shown that such correction is possible in DSMs.

Similar to facial expression, BMI is not mentioned in many stereophotogrammetry studies, especially those in genetic or developmental disorders. When it is considered, approaches have included scaling all face images to the same vertical size to assess shape only (Suttie et al., 2013). However, increasing BMI leads to increased lateral facial size, which is not accounted for using vertical scaling. Other studies mention

weight but suggest it is not relevant to the shape changes seen (Hennessy et al., 2010; Prasad et al., 2015). Other studies do not mention weight at all (Claes et al., 2014; Hammond et al., 2005; Kau et al., 2010; Paternoster et al., 2012). However, studies using DSMs should recognise that BMI can affect face shape significantly in terms of PCs, and data on height and weight should be collected for all participants. Another possible application of stereophotogrammetry and dense surface modelling is in the detection of certain craniofacial fat deposition patterns, as these patterns are thought to be a risk factor for obstructive sleep apnoea (Banabilh et al., 2009).

I have only performed one longitudinal study of subjects, in which the change in FSD over time was noted. The purpose was to study the effect of ageing, but given that the maximum time interval between images was less than three years, the study could only provide limited information to answer this question. However, the benefit of this longitudinal study was that it helps to confirm that FSD is a robust measure even as subjects grow older and their face images are compared to different age-matched reference face images to calculate FSD. An additional factor that was found to affect FSD in eight subjects was change in hair length or hairstyle between images, especially with inadvertent coverage of the forehead. Hair bands were used to avoid this.

## 9.6 FSD AND GENETIC DISORDERS INCLUDING SVS AND *HNF1B* MUTATIONS OR DELETIONS

In Chapter 8, I have shown that my protocol for image capture with this specific camera is suitable for analysing a specific genetic disorder, in this case disease due to *HNF1B* haploinsufficiency. I found that in a small group of eight affected subjects, there was an increase in FSD and also facial asymmetry. This study shows that FSD is useful in analysis of face shape in one disorder. If a disorder alters face shape from the average face, then it would be expected to increase FSD for affected patients. No common facial features were found though. It could be that FSD is sensitive to changes in face shape of

any kind, but that the identification of particular facial characteristics in a condition requires larger numbers of patients. Similar studies of FSD would be helpful in recurrent pathogenic SVs seen in epilepsy. For example, 15q13.3 microdeletions have been described to be associated with changes in the nasal tip or head size in a minority of patients (Lowther et al., 2015; van Bon et al., 2009). I was able to recruit five people with one recurrent SV (16p13.11 microdeletion) and no change in face shape was found, but this was a very small sample.

In Chapter 4, I showed that FSD is able to differentiate between faces of people with epilepsy and a pathogenic SV and those of people with epilepsy without a pathogenic SV. This was a two-step study in which a training set was used to calculate an optimum FSD threshold value in one DSM. In a second DSM, the training set was used to adjust the previous FSD value to an FSD threshold value for the second DSM. This FSD threshold was then tested on a new validation set. I obtained 66-80% sensitivity and 65-78% specificity using FSD of the whole face in detection of people with pathogenic SVs within a group with epilepsy.

FSD is therefore promising as a measurement of atypical face shape that may reflect underlying pathogenic copy number variation. However, there are a number of hurdles to overcome. My findings were in 181 subjects of whom 43 had pathogenic SVs. The study needs to be replicated in a larger sample. As mentioned above, recent research has suggested that ethnic differences within Europe can affect face shape, and some of my subjects were recruited at European sites or were recruited in the UK but had European ancestry. I explored the effect on FSD of potential face shape difference between people from the UK and people from other European centres in a sensitivity analysis, excluding people recruited from mainland Europe. I found that the changes in FSD remained significant. Also, the reference subjects were recruited both in the UK and mainland Europe, helping to minimise any differences due to different ethnic

origins within Europe. Not all subjects underwent BMI measurements and this may have affected results too. For 98 subjects who had BMI measurements, adjusting whole face FSD for BMI made it more sensitive and specific (78% and 75%, respectively). Of course, some SVs may affect BMI directly so correction for BMI may not be appropriate in SVs of unknown function. It was also seen that intellectual disability was more prevalent in people with pathogenic SVs, and intellectual disability was also associated with an altered face expression. Both intellectual disability and altered face expression were also excluded in a sensitivity analysis, and these are both likely to be potential confounders for future stereophotogrammetry studies of novel SVs too. All of these limitations need to be addressed by having a larger, more representative reference subject group and by more careful image acquisition and phenotyping. For example, it is possible with care to obtain adequate images in people with intellectual disability and children through distraction and repeated imaging, as has been used in my method and in others' guidelines (Heike et al., 2010).

After replication in a larger group of subjects with epilepsy, if FSD is successful, the next step would be to test it in subjects without epilepsy. It may be particularly useful in people with unexplained intellectual disability or other neurodevelopmental disorders for example. FSD would not replace genetic testing though. A finding of increased FSD could be an additional clinical reason to perform genetic testing in some individuals. Traditional clinical assessment of face shape is performed mainly by dysmorphologists (subspecialists within the field of clinical genetics), and I have shown that, even now, these specialists use only 'subjective' definitions for 62% of all terms describing the face and body (Section 1.1.4). There are only approximately 3 clinical geneticists per million population in the UK and 7% of clinical geneticists specialise in adults (Cooksey et al., 2006). Outside of clinical genetics, there is a lack of knowledge and a lack of confidence among other physicians, including neurologists (Burke et al., 2006; Mainous et al., 2013; Salm et al., 2014). Therefore, FSD may be able to assist

many clinicians in detection of atypical face shape. FSD itself does not appear to detect 'dysmorphic' features alone. In this group of subjects with pathogenic SVs, some had high FSD values but were not thought to be dysmorphic by their treating clinician. An exploratory study in which four neurology research fellows, who all conduct general neurology consultations in the UK, were asked to assess for dysmorphism also found that FSD of the whole face was more sensitive than three of the four neurologists in detecting pathogenic SVs. It would be useful to compare this with a trained dysmorphologist as well.

Another use of FSD is to further characterise the phenotype of known genetic disorders. An increase in FSD in a person could prompt more detailed face analysis, for example using heat maps to look objectively at different regions in the face (as shown in Chapter 8). Dense surface modelling and heat maps have already shown previously unrecognised facial features in mild Wolf-Hirschhorn syndrome and Bardet-Biedl syndrome (Hammond et al., 2012a; Tobin et al., 2008). This allows for better understanding of the genes affected in SVs or mutations. A genome-wide association study has shown that one gene, *PAX3*, expressed in the neural crest affects the shape of the nose (Paternoster et al., 2012). This was based on a landmark-based study of face shape. FSD lends itself well to similar genome-wide association studies to look for novel candidate genes involved in facial development. It has a clear advantage in being able to describe shape change in the whole face not just those parts that are able to be landmarked (Hammond and Suttie, 2012). This type of study would probably require recruitment of many thousands of subjects though, given that 4700 subjects were included in the earlier genome-wide association study.

In Chapter 5, I looked for any relationship of FSD to neurodevelopmental genes, and it appeared that FSD of the whole face was positively correlated with the number of genes highly expressed in the human embryonic forebrain. This was a crude approach

using a human transcriptome atlas. The finding can be explained from current knowledge of face and brain development, but would first need to be confirmed in another group as well. It rests on a number of assumptions, such as the level of gene expression in the forebrain being high for genes affecting facial development or that these genes all exert an equal effect on face shape at the embryonic stage. Therefore, further studies are needed, perhaps taking into account the growing knowledge of gene function, gene interactions and gene pathways.

An important practical limitation for clinical use of stereophotogrammetry in genetic disorders is the cost of the stereophotogrammetry system. The successor to the Canfield Vectra CR10 camera system, the three-camera pod Canfield M3 system, retails for \$13000 to \$27000 (in US dollars) depending on used or new systems. The clinical utility of the camera depends on its absolute cost, and also the cost relative to alternative diagnostic methods. Detection of SVs may be performed using whole genome sequencing or aCGH instead, for which the costs have been rapidly falling. The benefit of a stereophotogrammetry system may be to suggest those patients who would benefit from genetic testing. In addition, it has other clinical applications not studied here, such as for recording facial injuries or changes to the face due to anti-epileptic drug use.

## 9.7 REFLECTED FSD AND CORTICAL MALFORMATIONS INCLUDING HIPPOCAMPAL SCLEROSIS

Reflected FSD is a metric quantifying facial asymmetry that has been used in one previous study of face symmetry in autism spectrum disorders and found to be higher in boys with autism spectrum disorder and their mothers (Hammond et al., 2008). I found that, in people with epilepsy, reflected FSD was higher for people with hippocampal sclerosis and certain types of mixed malformations of cortical development. Therefore, reflected FSD appears to be able to detect abnormal

development of the central nervous system, as judged by the presence of malformations of cortical development or the presence of autism. This is in contrast to people with epilepsy and pathogenic SVs, for which there was no increase in reflected FSD. Hence, reflected FSD offers an adjunct to the measurement of FSD alone in detecting neurodevelopmental abnormalities. Reflected FSD has the added advantage of needing no reference group of people: each face surface image is compared to its own reflection in the vertical midline. This makes the measurement more robust with none of the limitations for FSD in terms of age or ethnicity of subjects. A set of reference subjects would still be helpful to understand more about facial asymmetry. For example, it is not clear if asymmetry is increased in people with more severe epilepsy due to trauma, and this could be a confounder if people with malformations of cortical development were more likely to have epilepsy of greater duration or greater severity than other people with epilepsy.

It does not appear that facial asymmetry is directly caused by mechanical effects of underlying brain lesions, given that lesions as small as hippocampal sclerosis, were also associated with significant facial asymmetry. Reflected FSD was highest in those with bilateral lesions, suggesting that face asymmetry may be a marker of the severity of underlying brain pathology in some people with epilepsy, rather than a marker of corresponding asymmetric brain pathology. This would fit with current theories of face symmetry as a phenotype for underlying 'developmental instability' in a person, in which greater sensitivity and exposure to genetic mutations, toxins, trauma or infections, may lead to greater facial asymmetry (Dongen, 2006). Related to this, it was found that left-handed people had greater facial asymmetry than right-handed people, and left-handedness has previously been associated with neurodevelopmental disorders (Lindell and Hudry, 2013).

Facial symmetry has been little studied in comparison to face shape, but with stereophotogrammetry and dense surface modelling, it is possible to investigate facial asymmetry at the same time as face image capture for atypical facial shape. Unlike for FSD and the presence of SVs, I was unable to replicate the results in a second validation set and this would be important to determine a threshold for reflected FSD which may predict the presence of a malformation of cortical development. A larger study set would also help explore the relationship of facial asymmetry with rarer malformations of cortical development, such as dysembryoplastic neuroepithelial tumours, which were not assessed individually here due to the small numbers of subjects.

## 9.8 LIMITATIONS OF THE PRESENT STUDY

### 9.8.1 RECRUITMENT OF REFERENCE SUBJECTS

As noted above (Section 9.4), my reference group limited analysis of FSD to white European people only up to an age of 52-54 years. For FSD to be used more widely, many more subjects need to be recruited into the reference group. This allows two benefits. Firstly, it means a greater number of faces may be 'averaged' limiting the contribution of any one face. Secondly, it allows finer comparisons. For example, it may be best to compare a 34-year-old man's face with an average of 30 faces of other 34-year-old men rather than 30 faces of those between 32 and 36 years. Also as noted, there are fine ethnic variations in face shape even in a category such as white Europeans (Kau et al., 2010). A larger reference group could match people to those of similar ethnic origins more accurately. No analysis at all was possible for people of Asian or African origin.

### 9.8.2 RECRUITMENT OF PEOPLE WITH EPILEPSY

A limitation in some of the individual studies here was the small sample size. The recruitment of a larger number of people is needed to explore all of my preliminary findings more fully. There are a number of factors to consider here. To recruit more

participants, more cameras are needed or the same camera will need to be used more times. Although the camera is portable, transporting it requires a specialised container over 1.5m long as well as transport of the calibration plate, personal computer and tripod separately. The camera also requires the use of an area of six square metres with privacy for subjects and needs re-calibration after set-up in the new location.

Portability was thus limited and the majority of participants were recruited at one location only (the Chalfont Centre for Epilepsy). This may have led to a bias in recruitment of those people who were attending clinic there or were admitted there. People in other parts of the country, people with well-controlled epilepsy not needing tertiary referral, or people with severe epilepsy who could not attend clinic, may all be under-represented.

I sought to recruit people who had been genotyped and especially those with SVs in one of the studies. Some genotyped subjects were found to have SVs as part of a previous study using SNP arrays (Heinzen et al., 2010). However, others were selected to have clinical array CGH, presumably due to clinical suspicion of the presence of an SV, which may have been due to facial dysmorphism. Therefore, I cannot generalise my findings of a more atypical face shape to all those people with epilepsy and SVs from the current study.

I have noted that full data did not exist for all participants, which was partly the reason for the number of subgroups in the studies above. Having data on neuropsychometry findings, imaging findings, genotyping, BMI and epilepsy or trauma history for every subject would be helpful. For example, body mass index was shown to affect face shape in a predictable manner and thus needs to be considered in any future study of face stereophotogrammetry. It also allows more powerful analysis: For example, it may be that only those subjects with intellectual disability show increased face asymmetry with a developmental brain lesion. However, I could not assess if this was the case as

not everyone who had undergone stereophotogrammetry and MRI scanning had had neuropsychological results.

### 9.8.3 STEREOPHOTOGRAMMETRY

Currently there are no formal guidelines on stereophotogrammetry for face shape analysis. However, I have shown that facial expression, face orientation and previous facial trauma may all affect findings. These may all have introduced errors in face shape analysis, especially angle of face orientation which could not be measured retrospectively. For subjects with altered face expression or those with facial trauma – both of which may be associated with severe epilepsy or intellectual disability – the only option was to exclude them in sensitivity analyses or include them as they were. It would be better if there was a method to ignore only the affected area of trauma.

Hair needs to be excluded specifically from studies too. Use of a hairband can help avoid this. Facial hair may also obscure or alter the face shape outline and the effect of this needs to be considered. Again, it would be helpful to be able to remove the affected area alone from analysis.

One other aspect is the part of the face included. In my study, the ears and the underside of the chin were not included. However, these are likely to provide useful information on face shape. Ear size and shape is useful in clinical dysmorphology and indeed there are more dysmorphological terms to describe the ear than any other part of the face (Table 1). Using a camera system that reliably captures the ears as well may increase the accuracy of the findings.

### 9.8.4 LANDMARKING

Landmarking is the major operator-dependent step between image capture and dense surface modelling. I have shown that there is a measurable intra-operator and inter-operator reproducibility error in landmark placement. This also appears to correlate

with experience of landmark placement and is complicated by the imprecise definition of many landmarks, which were inherited from direct anthropometry studies. I specifically assessed the effect of landmarking a second image of a control subject on FSD and found that landmarking could introduce an intra-operator error of 1-1.5% of FSD (Table 13). This error may be greater between operators and needs to be considered in future studies.

#### 9.8.5 DENSE SURFACE MODELLING AND FSD

I have noted that as part of the reduction in the volume of data, principal component analysis leads to loss of 1% of information about face shape. Given that principal component analysis seeks to explain as much variance as possible with limited PCs, it may be that the discarded 1% of information includes the most unique aspect of a subject's face. Again, this would blunt the accuracy of my technique and should be assessed using different thresholds, such as 98% or 99.9% of variance, to see if the findings are still valid.

### 9.9 CONCLUSIONS

The aims of my study were to show that stereophotogrammetry and dense surface modelling are reproducible and accurate techniques for face shape analysis as well as to show that structural variants and developmental brain lesions that are associated with epilepsy may affect face shape. Specifically, I expected to see changes in two new metrics of face shape, FSD (for atypical face shape) and reflected FSD (for face shape asymmetry). The findings may be summarised as:

1. Stereophotogrammetry and landmark placement using the Canfield Vectra CR10 camera is highly reproducible. Landmark placements and FSD are not affected by camera calibration but are affected by head position and face expression (Chapter 3).

2. In people with epilepsy, face shape, as measured by FSD, may be used to differentiate between those with pathogenic structural variants and those without such variants. Face shape may comprise the whole face or only a certain part of the face. Facial injury, facial expression, body mass index and anti-epileptic drug history do not appear to explain this difference (Chapter 4).
3. There may be a positive correlation between the number of genes, within a structural variant, which are highly expressed in the embryonic forebrain, and atypical face shape, as measured by FSD. FSD may also be able to detect people with epilepsy who have a deletion greater than 500 kilobases of DNA (Chapter 5).
4. Face shape varies in a predictable manner with changes in weight and body mass index. The effect of body mass index can be minimised in analyses using face shape. The method of face shape analysis used here is robust to changes in age in adults of up to 2.5 years (Chapter 6).
5. Facial asymmetry appears to be greater in people with epilepsy and a developmental brain lesion than those without a lesion; this includes just those people with hippocampal sclerosis. Bilateral lesions and facial injuries may increase facial asymmetry further (Chapter 7).
6. The method of analysis used here is also successful in investigating the effect on face shape and symmetry of a single genetic disorder (Chapter 8).

This thesis demonstrates a novel use for the techniques of stereophotogrammetry and dense surface modelling, in detecting atypical face shape by means of the metric 'FSD'. Factors affecting the reproducibility of FSD are assessed, including individual steps in the process of deriving FSD (such as camera calibration, image acquisition and image landmarking) as well as external factors, such as head positioning, facial expression and body mass index. I demonstrate methods to try and limit the effect of these potentially confounding variables. I then show that FSD appears to be able to detect face shape

change in a group of people with epilepsy and pathogenic structural variants, including an assessment of its accuracy as a diagnostic test. I also further study a metric of facial asymmetry, reflected FSD, that has been used before and show that it may be able to detect increased facial asymmetry in people with epilepsy and malformations of cortical development. Finally, I apply my techniques, including FSD and reflected FSD, to a group of people with a single genetic disorder and show that they appear to detect face shape changes that have not been previously described.

Stereophotogrammetry, dense surface modelling, FSD and reflected FSD appear to be promising tools in investigating children and adults with epilepsy and may also be useful in other populations which are associated with genetic disorders, including structural variants, or neurodevelopmental disorders. The techniques are more objective than previous methods of face shape measurement and show high reproducibility.

## 9.10 FUTURE DIRECTIONS

### 9.10.1 STEREOPHOTOGRAMMETRY

Stereophotogrammetry cameras continue to evolve incrementally, but the basic premise remains the same. An example is a new 3dMD camera system that allows for a series of 3D face images to be acquired at a rate of 60 images per second. This is to allow the operator to choose the most appropriate image for further analysis. From my study, it is important to choose an image with a neutral expression and natural head position with the eyes open and mouth closed. As noted, children and people with intellectual disability may be less able to do this for a single image and thus being able to acquire images rapidly is helpful. Another manufacturer, Canfield, has developed a handheld 3D camera system which has now been validated for face imaging (Camison et al., 2017). Greater portability allows for easier recruitment of subjects. Cost remains a major barrier for most of these professional imaging systems. One team has

experimented with using much cheaper smartphones for 3D image capture, though there was decreased accuracy and the need for marking of the subject's face (Barbero-Garcia et al., 2017).

Landmarking is the only operator-dependent step after image acquisition and requires clinicians or researchers to be trained in identification of face landmarks and even with training, inter-operator reproducibility is worse than intra-operator reproducibility in my study and others (Gwilliam et al., 2006; Plooij et al., 2009; Schaaf et al., 2010).

Therefore, avoiding this step or making it objective would be helpful in reducing error in any face shape analysis. There are now techniques being developed for such automated landmark placement, which may even work on incomplete face surfaces (Katina et al., 2016; Sukno et al., 2015).

#### 9.10.2 DENSE SURFACE MODELLING AND FSD

Dense surface modelling remains the most sophisticated method of analysing face shape but improvements are possible for future studies. One current limitation is the need to create a separate DSM for each analysis. Separate models were used to analyse each region of the face: the whole face, the periorbital region, the perinasal region and the perioral region. Another separate model was used to analyse asymmetry, which had involved the creation of 3D mirror images of each subject's face image. The creation of dynamic models in which a new face image may be added to a DSM database of existing face images, without recalculating the principal components for every face image again, would significantly speed up analysis and would be necessary for routine clinical use. This would entail new software programming as principal component analysis may need to be replaced by an alternative method, such as factor analysis, with adjustment of the metric then used for face shape.

After principal component calculation, I used a metric which calculates the difference between an 'average' matched face and the subject's face across all principal

components. Each principal component is equally weighted in this calculation. However we know that later principal components will explain a smaller amount of variation. It could be argued that the differences in principal components should also be weighted. On the other hand, later principal components may be the ones that are best at identifying dysmorphism and simply occur later because they are the ones that are deviated from average more rarely in the population. Therefore, in a larger genotyped dataset, it would be useful to assess the accuracy of different metrics based on principal components, of which FSD is just one. An algorithm could also be used either within the DSM software or in an external application to automate the calculation of FSD and reflected FSD.

### 9.10.3 GENETIC STUDIES

The studies here suggest that face shape is affected in a number of pathogenic SVs thought to cause epilepsy. It would be useful to assess if this is true for all known pathogenic SVs in people with epilepsy or if there are only certain SVs that affect face shape. There are now more sophisticated classifications of SVs and pathogenicity may be based on larger databases of known SVs as well as SVs containing epilepsy genes or in known epilepsy hotspots (Lal et al., 2015; Fry et al., 2016).

Related to this, further studies of individual SVs are needed. I found no common face shape features in five people with 16p13.11 deletions, the largest group in my study for an individual SV. Larger numbers for this and other SVs are needed to detect more subtle effects on face shape. Even if changes are identified, studies in different ethnicities would be useful to see if similar effects are noted on face shape.

Apart from pathogenic SVs, other types of SVs and other genetic mutations may also affect face shape. Careful face shape phenotyping, using stereophotogrammetry, in these cases may help understand the effect of these genetic changes.

For some adults with particular genetic syndromes, stereophotogrammetry and DSM can quantify the change in facial shape from the 'average' face as well as showing the effect on face shape graphically. This graphical representation may help clinicians, especially those in adult medicine without dysmorphology training, to recognise these features in future. Serial stereophotogrammetry could further identify the age at which a condition exerts the most prominent effect on face shape.

There has been one genome-wide association study using facial landmarks (Paternoster et al., 2012) which found an association between nasion-menton distance and the *PAX3* gene. Given that principal components and FSD incorporate information about face shape outside of the landmarks themselves, genome-wide association studies using either of these parameters may find genes associated with face shape more successfully.

Further studies of genetic disorders could look at children primarily. Older children would be able to co-operate with stereophotogrammetry but they would be hopefully less exposed to environmental factors that may affect face shape such as drugs or trauma. Also, some genetic disorders are known to be less easily detected after adulthood (Hammond et al., 2005).

#### 9.10.4 IMAGING STUDIES

The most logical follow-up study after my findings in Chapter 7 would be to confirm if all types of developmental brain lesion are associated with increased facial asymmetry. I had only shown this for people with hippocampal sclerosis. Also, it would be useful to know if the presence of bilateral lesions increases asymmetry for lesions other than hippocampal sclerosis. Handedness appears to exert a significant effect on asymmetry with increased facial asymmetry seen in left-handed people. It is unclear what the mechanism for this is and it may be from confounding variables such as the known higher prevalence of left-handedness in some neurodevelopmental disorders. Detailed

phenotyping, including neuropsychological assessment and full clinical history, can help detect such confounders.

I was only able to perform an assessment of face shape in relation to the categorical variables of presence or absence of unilateral and bilateral developmental brain lesions, as diagnosed by a neurologist or neuroradiologist. Co-registration of MRI scans with 3D face images would be very helpful to quantitatively assess face shape with intracranial lesions, although this has not been performed with stereophotogrammetry yet. One difficulty in combining the two sets of data is that MRI provides volumetric data whereas stereophotogrammetry provides a 'point cloud' of sets of 3D co-ordinates only. MRI scanners are able to capture face shape information themselves and it may be simpler to use just this modality. One group used MRI scans with software processing afterwards to produce a face surface image with a 'point cloud' and showed that 86% of data points were within 2mm of the gold standard of stereophotogrammetry (Knoops et al., 2017). This may be sufficient accuracy to have accurate co-localisation with brain imaging findings.

For malformations of cortical development, the classifications have changed since the diagnosis was made in some subjects (Barkovich et al., 2012). Therefore, it may be that future studies with the new classification are better able to detect an association between face asymmetry and imaging findings. As well as the classification, MRI scanners and imaging protocols are more advanced. Scanners with magnetic field strengths of 7 Tesla have entered clinical practice and been shown to detect lesions, particularly subtle FCD, not seen in lower field strength scanners (de Ciantis et al., 2016).

## APPENDIX 1: IMAGE CAPTURE PROTOCOL

An image capture protocol for other researchers in the group was created by the author to enable image capture by them as well. This protocol contains practical information on the use of the Canfield Vectra CR10 camera, the Mirror Vectra software and basic image quality control procedures at the time of image capture. It is reproduced overleaf.

## PROTOCOL FOR IMAGE CAPTURE

### SETTING UP

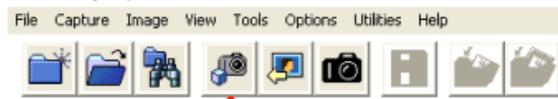
- Access Room 8 aka Neurogenetics room
  - Keys kept by XXXX
  - Code known to above
- Switch on equipment
  - Uncover lenses on both sides of the camera
  - Switch on PC and both monitors
  - Switch on camera
    - Press red switch behind device
    - Press black switch above & to the side of the red switch
    - Two green lights should appear above the black switch now
    - Check there is one green light lit on the black box suspended behind the camera
    - If the camera is not recognised or switched on after the program is open, press the button on the black box to synchronise it again
- Log on to the PC
  - One Windows login used for all users (XXXX)
  - Emanuele and I know the password currently
- Open Mirror Vectra CR10
  - Click on the desktop icon
- Check the patient has consented (using Epigen version 7 forms)
  - Black box labelled DSM has a red folder with consent and blood forms
- Keep a photocopy of signed consent forms in the lever arch A4 file marked DSM

### CAMERA POSITIONING

- The camera should have an upward tilt of  $\sim 20^\circ$  (so lenses point forward and up)
- The tripod has screws halfway along each leg that can be loosened to change the angle and length of the legs
  - If this is done, check the spirit level at the top of the tripod
  - The tripod mount should be horizontal after adjustments
- Use the rotating handle below the grey monitor to change the camera height
- Change the seat height wherever possible instead

### CALIBRATION

- The camera should be calibrated for every session
- This involves capturing 2 images whilst holding up a black plate to the camera
- After calibration, apart from the height of the camera, no other settings should be changed, including position and lighting
- From the main window, open any patient record by double-clicking on it
- The Patient Chart screen now appears
- Click the 3D Capture icon on the top menu bar



- Click the Calibrate icon



- Hold the black plate with the 'L' facing the camera
- The 'L' should face the camera and be positioned in the middle of both views
- The plate should be angled UP at 45°
- Click Capture Images in the dialog box
- Use the remote mouse pointer to press it if operating the camera alone
- Now angle the plate DOWN at 45°
- Press Capture Images to take the second calibration image
- Wait for the software to analyse the images
- A message will indicate if calibration was successful or not
- If unsuccessful, repeat the above, and pay attention to the plate positioning
- If successful, click on File → Exit to return to the main screen

#### IMAGE CAPTURE OF A NEW PATIENT

- These are all the steps in capturing images for a patient
- **Create a new record**
  - Start from the Main Screen
  - Click on the top left icon on the menu bar or go to File → New Patient



- Enter the following details into the box that opens
  - Surname
  - First name
  - Date of birth
  - MRN (use their hospital number here)
    - For relatives, use the hospital number of the patient
  - Sex
  - Race
- A new Patient Chart screen now opens

- **Position patient & set up for photo**

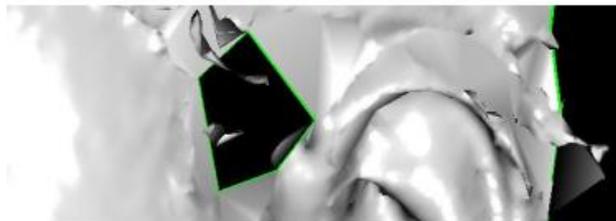


- Wait for the camera to sync with the PC and display two views in the 3D Capture screen

- Remove glasses, head coverings, etc
- Loosen collar to expose chin adequately
- Uncover forehead and ears with hairband and clips if needed
  - Hair in front of ears needs to be clipped above or flattened
  - Black box labelled DSM has hairband and clips
- Tilt the face up slightly
  - Up enough so that the chin is seen
  - Not too high so that the skin is stretched or the parts of the face are not visible (eg. bridge of nose – see below)
- Eyes are open and looking straight ahead between the two lenses
- Look neutral with lips together, and not smiling
- **Image capture**
  - Click Capture when the patient is ready and is still



- Wait for 1s after the flash; image is then captured
- Wait approx. 1 minute for the 3D image to be created
- Do not interrupt this process unless the image was obviously not adequate and needs repeating. Interrupting the 3D image creation means the image will be lost.
- The Mirror 3D Analysis screen will then open to display the 3D image
- Rotate the image to look for any defects. The forehead, tragus (both) and jaw should be included



*Example of an absolute defect. This is due to hair in front of the ear, but is not an issue as it is above the tragus and the ear is still captured.*

- Switch to wireframe mode to eliminate the surface textures, by clicking on the grey face icon



- Look for subtle areas of 'filling in' by the software due to inadequate views



*Example of 'filling in' just over the bridge of the nose. The face was tilted too high so this part of the nose was not seen by the camera and the software filled in the defect with a flat surface.*

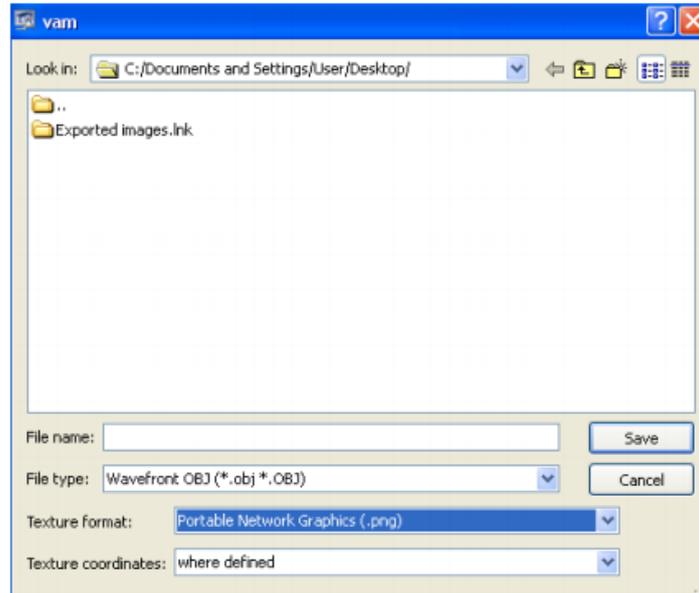
- If the image is inadequate, close the 3D Analysis screen and repeat above
- **Add image details**
  - Close the 3D Analysis screen to return to the Patient Chart screen
  - Select all of the images taken by clicking on each of them once



- Click Edit Data on the right-hand menu
- Select Attribute on the right-hand menu and click on Relative.
- From the drop-down list, select one of the options: Patient (if the person is a patient and not a relative), Sister, Brother, Father, Mother
- Select Physician in the same menu, then select your initials

#### **EXPORTING IMAGES**

- Image processing and analysis is done using proprietary software at the Institute of Child Health
- To export images in a suitable format for this, do the following:
- From the Main Screen, double-click a patient record to open it
- From the Patient Chart screen, double-click a particular image to open it
- In the Mirror 3D Analysis screen, select File → Export
- Choose File type as Wavefront OBJ (\*.obj)
- Choose Texture format as Portable Network Graphics (\*.png)



- Save the file
  - This creates 3 files: xxx.mtl, xxx.obj, xxx.png
  - The .obj & .png files are essential for image processing

#### **CLOSING DOWN**

- Tidy up everything!
- Close Mirror Vectra, then turn off the PC
- Turn off both the red and black switches on the camera
- Cover the camera lenses
- Lock the door!

## REFERENCES

- Adameyko, I., Fried, K., 2016. The nervous system orchestrates and integrates craniofacial development: a review. *Front Physiol.* 7:49.
- Akgül, A.A., Toygar, T.U., 2002. Natural craniofacial changes in the third decade of life: a longitudinal study. *Am J Orthod Dentofacial Orthop.* 122:512–522.
- Albert, A.M., Ricanek, K., Jr, Patterson, E., 2007. A review of the literature on the aging adult skull and face: implications for forensic science research and applications. *Forensic Sci. Int.* 172:1–9.
- Albertz, J., 2007. A look back - 140 years of photogrammetry. *Photogrammetric Engineering and Remote Sensing.* 73:504–6.
- Aldridge, K., Boyadjiev, S.A., Capone, G.T., DeLeon, V.B., Richtsmeier, J.T., 2005. Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *Am J Med Genet A* 138A:247–253.
- Alkan, C., Coe, B.P., Eichler, E.E., 2011. Genome structural variation discovery and genotyping. *Nat Rev Genet* 12:363–376.
- Allanson, J.E., 1997. Objective techniques for craniofacial assessment: What are the choices? *Am J Med Genet* 70:1–5.
- Allanson, J.E., Biesecker, L.G., Carey, J.C., Hennekam, R.C.M., 2009a. Elements of morphology: Introduction. *Am J Med Genet A* 149A:2–5.
- Allanson, J.E., Cunniff, C., Hoyme, H.E., McGaughran, J., Muenke, M., Neri, G., 2009b. Elements of morphology: Standard terminology for the head and face. *Am J Med Genet A* 149A:6–28.
- Allanson, J.E., O'Hara, P., Farkas, L.G., Nair, R.C., 1993. Anthropometric craniofacial pattern profiles in Down syndrome. *Am J Med Genet* 47:748–752.
- Almukhtar, A., Ju, X., Khambay, B., McDonald, J., Ayoub, A., 2014. Comparison of the accuracy of voxel based registration and surface based registration for 3D assessment of surgical change following orthognathic surgery. *PLoS ONE* 9:e93402.
- Antonarakis, S.E., Lyle, R., Dermitzakis, E.T., Reymond, A., Deutsch, S., 2004. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat Rev Genet* 5:725–738.
- Aristotle, 1913. *Physiognomonica*, in: Loveday, T., Forster, E. (Trans.), *Opuscula, The Works of Aristotle*. Clarendon Press, Oxford.
- Aronica, E., Crino, P.B., 2014. Epilepsy related to developmental tumors and malformations of cortical development. *Neurotherapeutics* 11:251–268.
- Artopoulos, A., Buytaert, J.N., Dirckx, J.J.J., Coward, T.J., 2014. Comparison of the accuracy of digital stereophotogrammetry and projection moiré profilometry for three-dimensional imaging of the face. *Int J Oral Maxillofac Surg.* 43:654–662.

- Ashwell, M., Gunn, P., Gibson, S., 2012. Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis. *Obes Rev* 13:275–286.
- Aung, S.C., Ngim, R.C., Lee, S.T., 1995. Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements. *Br J Plast Surg* 48:551–558.
- Ayoub, A., Garrahy, A., Hood, C., White, J., Bock, M., Siebert, J.P., et al., 2003. Validation of a vision-based, three-dimensional facial imaging system. *Cleft Palate Craniofac. J.* 40:523–529.
- Bahi-Buisson, N., Ville, D., Eisermann, M., Plouin, P., Kaminska, A., Chiron, C., 2005. [Epilepsy in chromosome aberrations]. *Arch Pediatr* 12:449–458.
- Banabilh, S.M., Suzina, A.H., Dinsuhaimi, S., Samsudin, A.R., Singh, G.D., 2009. Craniofacial obesity in patients with obstructive sleep apnea. *Sleep Breath* 13:19–24.
- Barbero-Garcia, I., Lerma, J.L., Marqués-Mateu, Á., Miranda, P., 2017. Low-cost smartphone-based photogrammetry for the analysis of cranial deformation in infants. *World Neurosurg.* 102:545-554.
- Barkovich, A.J., Guerrini, R., Kuzniecky, R.I., Jackson, G.D., Dobyns, W.B., 2012. A developmental and genetic classification for malformations of cortical development: update 2012. *Brain* 135:1348–1369.
- Bartzela, T., Katsaros, C., Rønning, E., Rizell, S., Semb, G., Bronkhorst, E., et al., 2012. A longitudinal three-center study of craniofacial morphology at 6 and 12 years of age in patients with complete bilateral cleft lip and palate. *Clin Oral Investig* 16:1313–1324.
- Beard, L.F., Burke, P.H., 1967. Evolution of a system of stereophotogrammetry for the study of facial morphology. *Med Biol Illus* 17:20–25.
- Bellanné-Chantelot, C., Chauveau, D., Gautier, J.-F., Dubois-Laforgue, D., Clauin, S., Beaufiles, S., et al., 2004. Clinical spectrum associated with hepatocyte nuclear factor-1beta mutations. *Ann Intern Med* 140:510–517.
- Ben-Menachem, E., 2007. Weight issues for people with epilepsy--a review. *Epilepsia* 48 Suppl 9:42–45.
- Ben-Salem, S., Al-Shamsi, A.M., John, A., Ali, B.R., Al-Gazali, L., 2015. A novel whole exon deletion in *WWOX* gene causes early epilepsy, intellectual disability and optic atrophy. *J Mol Neurosci* 56:17–23.
- Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, J.H., van Emde Boas, W., et al., 2010. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia* 51:676–685.
- Bhuiyan, Z.A., Klein, M., Hammond, P., van Haeringen, A., Mannens, M.M.M., Van Berckelaer-Onnes, I., et al., 2006. Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. *J Med Genet* 43:568–575.
- Biesecker, L.G., 2005. Mapping phenotypes to language: a proposal to organize and standardize the clinical descriptions of malformations. *Clin Genet* 68:320–326.

Biesecker, L.G., Aase, J.M., Clericuzio, C., Gurrieri, F., Temple, I.K., Toriello, H., 2009. Elements of morphology: Standard terminology for the hands and feet. *Am J Med Genet A* 149A:93–127.

Bockenbauer, D., Jaureguiberry, G., 2016. HNF1B-associated clinical phenotypes: the kidney and beyond. *Pediatr Nephrol* 31:707–714.

Brunklaus, A., Ellis, R., Reavey, E., Forbes, G.H., Zuberi, S.M., 2012. Prognostic, clinical and demographic features in SCN1A mutation-positive Dravet syndrome. *Brain* 135:2329–2336.

Brunklaus, A., Zuberi, S.M., 2014. Dravet syndrome--from epileptic encephalopathy to channelopathy. *Epilepsia* 55:979–984.

Buckland, P.R., 2003. Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Ann Med* 35:308–315.

Buckley, P.F., Dean, D., Bookstein, F.L., Han, S., Yerukhimovich, M., Min, K.-J., et al., 2005. A three-dimensional morphometric study of craniofacial shape in schizophrenia. *Am J Psychiatry* 162:606–608.

Burke, S., Stone, A., Bedward, J., Thomas, H., Farndon, P., 2006. A “neglected part of the curriculum” or “of limited use”? Views on genetics training by nongenetics medical trainees and implications for delivery. *Genet Med* 8:109–115.

Calhoun, W.J., 2003. Nocturnal asthma. *Chest* 123:399S–405S.

Caminiti, C.B., Hesdorffer, D.C., Shostak, S., Goldsmith, J., Sorge, S.T., Winawer, M.R., et al., 2016. Parents’ interest in genetic testing of their offspring in multiplex epilepsy families. *Epilepsia* 57:279–287.

Camison, L., Bykowski, M., Lee, W.W., Carlson, J.C., Roosenboom, J., Goldstein, J.A., et al., 2017. Validation of the Vectra H1 portable three-dimensional photogrammetry system for facial imaging. *Int J Oral Maxillofac Surg* S0901-5027:31606-5.

Carey, J.C., Cohen, M.M., Curry, C.J.R., Devriendt, K., Holmes, L.B., Verloes, A., 2009. Elements of morphology: Standard terminology for the lips, mouth, and oral region. *Am J Med Genet A* 149A:77–92.

Caron, L., Tihy, F., Dallaire, L., 1999. Frequencies of chromosomal abnormalities at amniocentesis: over 20 years of cytogenetic analyses in one laboratory. *Am J Med Genet.* 82:149–154.

Cassidy, S.B., Schwartz, S., Miller, J.L., Driscoll, D.J., 2012. Prader-Willi syndrome. *Genet Med* 14:10–26.

Catarino, C.B., Kasperavičiūtė, D., Thom, M., Cavalleri, G.L., Martinian, L., Heinzen, E.L., et al., 2011. Genomic microdeletions associated with epilepsy: not a contraindication to resective surgery. *Epilepsia* 52:1388–1392.

Chakravarty, M.M., Aleong, R., Leonard, G., Perron, M., Pike, G.B., Richer, L., et al., 2011. Automated analysis of craniofacial morphology using magnetic resonance images. *PLoS One* 6:e20241.

- Chen, B., Choi, H., Hirsch, L.J., Moeller, J., Javed, A., Kato, K., et al., 2015. Cosmetic side effects of antiepileptic drugs in adults with epilepsy. *Epilepsy Behav* 42:129–137.
- Chen, Z.-C., Albdour, M.N., Lizardo, J.A., Chen, Y.-A., Chen, P.K.-T., 2015. Precision of three-dimensional stereo-photogrammetry (3dMD™) in anthropometry of the auricle and its application in microtia reconstruction. *J Plast Reconstr Aesthet Surg* 68:622–631.
- Chiba, M., Terazawa, K., 1998. Estimation of stature from somatometry of skull. *Forensic Sci Int* 97:87–92.
- Chiu, C.S., Clark, R.K., 1991. Reproducibility of natural head position. *J Dent* 19:130–131.
- Christianson, C.A., McWalter, K.M., Warren, N.S., 2005. Assessment of allied health graduates' preparation to integrate genetic knowledge and skills into clinical practice. *J Allied Health* 34:138–144.
- Claes, P., Liberton, D.K., Daniels, K., Rosana, K.M., Quillen, E.E., Pearson, L.N., et al., 2014. Modeling 3D facial shape from DNA. *PLoS Genet* 10:e1004224.
- Coffinier, C., Gresh, L., Fiette, L., Tronche, F., Schütz, G., Babinet, C., et al., 2002. Bile system morphogenesis defects and liver dysfunction upon targeted deletion of HNF1beta. *Development* 129:1829–1838.
- Coffinier, C., Thépot, D., Babinet, C., Yaniv, M., Barra, J., 1999. Essential role for the homeoprotein vHNF1/HNF1beta in visceral endoderm differentiation. *Development* 126:4785–4794.
- Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., et al., 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 35:2013–2025.
- Collaborative Group for Epidemiology of Epilepsy, 1988. Adverse reactions to antiepileptic drugs: a follow-up study of 355 patients with chronic antiepileptic drug treatment. *Collaborative Group for Epidemiology of Epilepsy. Epilepsia* 29:787–793.
- Cooksey, J.A., Forte, G., Flanagan, P.A., Benkendorf, J., Blitzer, M.G., 2006. The medical genetics workforce: an analysis of clinical geneticist subgroups. *Genet Med* 8:603–614.
- Coppola, A., Chinthapalli, K., Hammond, P., Sander, J.W., Sisodiya, S.M., 2013. Pediatric diagnosis not made until adulthood: a case of Wolf-Hirschhorn syndrome. *Gene* 512:532–535.
- Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., Helms, J.A., 2011. Cranial neural crest cells on the move: Their roles in craniofacial development. *Am J Med Genet* 155:270–279.
- Corey, L.A., Pellock, J.M., Kjeldsen, M.J., Nakken, K.O., 2011. Importance of genetic factors in the occurrence of epilepsy syndrome type: a twin study. *Epilepsy Res* 97:103–111.
- Coussens, A.K., van Daal, A., 2005. Linkage disequilibrium analysis identifies an FGFR1 haplotype-tag SNP associated with normal variation in craniofacial shape. *Genomics* 85:563–573.

- Cox-Brinkman, J., Vedder, A., Hollak, C., Richfield, L., Mehta, A., Orteu, K., et al., 2007. Three-dimensional face shape in Fabry disease. *Eur J Hum Genet* 15:535–542.
- Curtis, C., Lynch, A.G., Dunning, M.J., Spiteri, I., Marioni, J.C., Hadfield, J., et al., 2009. The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* 10:588.
- D'Angelo, A., Bluteau, O., Garcia-Gonzalez, M.A., Gresh, L., Doyen, A., Garbay, S., et al., 2010. Hepatocyte nuclear factor 1alpha and beta control terminal differentiation and cell fate commitment in the gut epithelium. *Development* 137:1573–1582.
- De Ciantis, A., Barba, C., Tassi, L., Cosottini, M., Tosetti, M., Costagli, M., et al., 2016. 7T MRI in focal epilepsy with unrevealing conventional field strength imaging. *Epilepsia* 57:445-54.
- de la Torre-Ubieta, L., Won, H., Stein, J.L., Geschwind, D.H., 2016. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* 22:345–361.
- de Menezes, M., Rosati, R., Allievi, C., Sforza, C., 2009. A photographic system for the three-dimensional study of facial morphology. *Angle Orthod* 79:1070–1077.
- de Menezes, M., Rosati, R., Ferrario, V.F., Sforza, C., 2010. Accuracy and reproducibility of a 3-dimensional stereophotogrammetric imaging system. *J Oral Maxillofac Surg* 68:2129–2135.
- de Smith A.J., Walters R.G., Froguel P., Blakemore A.I., 2008. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenet Genome Res.* 123:17-26.
- de Souza, M.A., McAllister, C., Suttie, M., Perrotta, C., Mattina, T., Faravelli, F., et al., 2013. Growth hormone, gender and face shape in Prader-Willi syndrome. *Am J Med Genet A* 161A:2453–2463.
- Dellinger, A.E., Saw, S.-M., Goh, L.K., Seielstad, M., Young, T.L., Li, Y.-J., 2010. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res.* 38:e105.
- Demyer, W., Zeman, W., Palmer, C.G., 1964. The face predicts the brain: Diagnostic significance of median facial anomalies for holoprosencephaly (arhinencephaly). *Pediatrics* 34:256–263.
- Diliberti, J.H., 1988. Use of computers in dysmorphology. *J Med Genet* 25:445–453.
- Diliberti, J.H., Olson, D.P., 1991. Photogrammetric evaluation in clinical genetics: theoretical considerations and experimental results. *Am J Med Genet* 39:161–166.
- Dindaroğlu, F., Kutlu, P., Duran, G.S., Görgülü, S., Aslan, E., 2015. Accuracy and reliability of 3D stereophotogrammetry: A comparison to direct anthropometry and 2D photogrammetry. *Angle Orthod.* 86:487-494.
- Dixit, A., Suri, M., 2016. When the face says it all: dysmorphology in identifying syndromic causes of epilepsy. *Pract Neurol* 16:111–121.

- Djordjevic, J., Pirttiniemi, P., Harila, V., Heikkinen, T., Toma, A.M., Zhurov, A.I., et al., 2013. Three-dimensional longitudinal assessment of facial symmetry in adolescents. *Eur J Orthod* 35(2):143-51.
- Dongen, S.V., 2006. Fluctuating asymmetry and developmental instability in evolutionary biology: past, present and future. *J Evol Biol* 19:1727-1743.
- Douglas, M.S., 1986. Moire topography of the face: Its application in measuring soft tissue asymmetries. *American Journal of Orthodontics* 89:177.
- Eadie, M.J., 2006. The epileptology of William Aldren Turner. *J Clin Neurosci* 13:9-13.
- Edghill, E.L., Bingham, C., Ellard, S., Hattersley, A.T., 2006. Mutations in hepatocyte nuclear factor-1beta and their related phenotypes. *J Med Genet* 43:84-90.
- Edghill, E.L., Oram, R.A., Owens, M., Stals, K.L., Harries, L.W., Hattersley, A.T., et al., 2008. Hepatocyte nuclear factor-1beta gene deletions--a common cause of renal disease. *Nephrol Dial Transplant* 23:627-635.
- Eling, P., Draaisma, D., Conradi, M., 2011. Gall's visit to The Netherlands. *J Hist Neurosci* 20:135-150.
- Elsea, S.H., Girirajan, S., 2008. Smith-Magenis syndrome. *Eur J Hum Genet* 16:412-421.
- Evison, M., Dryden, I., Fieller, N., Mallett, X., Morecroft, L., Schofield, D., et al., 2010. Key parameters of face shape variation in 3D in a large sample. *J Forensic Sci* 55:159-162.
- Fagertun, J., Harder, S., Rosengren, A., Moeller, C., Werge, T., Paulsen, R.R., et al., 2014. 3D facial landmarks: Inter-operator variability of manual annotation. *BMC Med Imaging* 14:35.
- Falconer, M.A., Davidson, S., 1973. Coarse features in epilepsy as a consequence of anticonvulsant therapy. Report of cases in two pairs of identical twins. *Lancet* 2:1112-1114.
- Farkas, L., 1994. *Anthropometry of the head and face*, 2nd ed. ed. Raven Press, New York.
- Farkas, L.G., Cheung, G., 1981. Facial asymmetry in healthy North American Caucasians. An anthropometrical study. *Angle Orthod* 51:70-77.
- Farkas, L.G., Deutsch, C.K., 1996. Anthropometric determination of craniofacial morphology. *Am J Med Genet* 65:1-4.
- Farkas, L.G., Eiben, O.G., Sivkov, S., Tompson, B., Katic, M.J., Forrest, C.R., 2004. Anthropometric measurements of the facial framework in adulthood: age-related changes in eight age categories in 600 healthy white North Americans of European ancestry from 16 to 90 years of age. *J Craniofac Surg* 15:288-298.
- Farkas, L.G., Hreczko, T.A., Kolar, J.C., Munro, I.R., 1985. Vertical and horizontal proportions of the face in young adult North American Caucasians: revision of neoclassical canons. *Plast Reconstr Surg* 75:328-338.

- Farkas, L.G., Katic, M.J., Forrest, C.R., Litsas, L., 2001. Surface anatomy of the face in Down's syndrome: linear and angular measurements in the craniofacial regions. *J Craniofac Surg* 12:373–380.
- Farooqi, I.S., O'Rahilly, S., 2005. New advances in the genetics of early onset obesity. *Int J Obes* 29:1149–1152.
- Federation of the Royal Colleges of Physicians of the UK, 2011. Census of consultant physicians and medical registrars in the UK, 2010: data and commentary.
- Feenstra, B., Pasternak, B., Geller, F., Carstensen, L., Wang, T., Huang, F., et al., 2014. Common variants associated with general and MMR vaccine-related febrile seizures. *Nat Genet* 46:1274–1282.
- Feingold, M., Bossert, W.H., 1974. Normal values for selected physical parameters: an aid to syndrome delineation. *Birth Defects* 10:1–16.
- Ferrario, V.F., Dellavia, C., Colombo, A., Sforza, C., 2004. Three-dimensional assessment of nose and lip morphology in subjects with down syndrome. *Ann Plast Surg* 53:577–583.
- Ferrario, V.F., Sforza, C., Poggio, C.E., Cova, M., Tartaglia, G., 1998. Preliminary evaluation of an electromagnetic three-dimensional digitizer in facial anthropometry. *Cleft Palate Craniofac. J.* 35:9–15.
- Ferrario, V.F., Sforza, C., Poggio, C.E., Schmitz, J.H., 1999. Soft-tissue facial morphometry from 6 years to adulthood: a three-dimensional growth study using a new modeling. *Plast Reconstr Surg* 103:768–778.
- Fisher, R.S., van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., et al., 2005. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* 46:470–472.
- Fisher, R.S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J.H., Elger, C.E. et al., 2014. ILAE official report: a practical clinical definition of epilepsy. *Epilepsia*. 55:475-82.
- Flegal, K.M., Kit, B.K., Orpana, H., Graubard, B.I., 2013. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA* 309:71–82.
- Forsberg, C.M., 1979. Facial morphology and ageing: a longitudinal cephalometric investigation of young adults. *Eur J Orthod* 1:15–23.
- Fourie, Z., Damstra, J., Gerrits, P.O., Ren, Y., 2011. Evaluation of anthropometric accuracy and reliability using different three-dimensional scanning systems. *Forensic Sci Int* 207:127–134.
- Fry, A.E., Rees, E., Thompson, R., Mantripragada, K., Blake, P., Jones, G., et al., 2016. Pathogenic copy number variants and SCN1A mutations in patients with intellectual disability and childhood-onset epilepsy. *BMC Med Genet* 17:34.
- Gaitatzis, A., Carroll, K., Majeed, A., W Sander, J., 2004. The epidemiology of the comorbidity of epilepsy in the general population. *Epilepsia* 45:1613–1622.

- Galantucci, L.M., Di Gioia, E., Lavecchia, F., Percoco, G., 2014. Is principal component analysis an effective tool to predict face attractiveness? A contribution based on real 3D faces of highly selected attractive women, scanned with stereophotogrammetry. *Med Biol Eng Comput* 52:475–489.
- Galizia, E.C., Srikantha, M., Palmer, R., Waters, J.J., Lench, N., Ogilvie, C.M., et al., 2012. Array comparative genomic hybridization: results from an adult population with drug-resistant epilepsy and co-morbidities. *Eur J Med Genet* 55:342–348.
- Georgala, P.A., Carr, C.B., Price, D.J., 2011. The role of Pax6 in forebrain development. *Dev Neurobiol* 71:690–709.
- Gerth, D.J., 2015. Structural and volumetric changes in the aging face. *Facial Plast Surg* 31:3–9.
- Ghoddousi, H., Edler, R., Haers, P., Wertheim, D., Greenhill, D., 2007. Comparison of three methods of facial measurement. *Int J Oral Maxillofac Surg* 36:250–258.
- Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., et al., 2012. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med* 367:1321–1331.
- Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., et al., 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 42:203–209.
- Goodman, R.M., 1977. *Atlas of the face in genetic disorders*, 2d ed. ed. Mosby, Saint Louis.
- Goulding, M.D., Chalepakis, G., Deutsch, U., Erselius, J.R., Gruss, P., 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J* 10:1135–1147.
- Greshock, J., Feng, B., Nogueira, C., Ivanova, E., Perna, I., Nathanson, K., et al., 2007. A comparison of DNA copy number profiling platforms. *Cancer Res.* 67:10173–10180.
- Gripp, K.W., Baker, L., Telegrafi, A., Monaghan, K.G., 2016. The role of objective facial analysis using FDNA in making diagnoses following whole exome analysis. Report of two patients with mutations in the BAF complex genes. *Am J Med Genet A* 170:1754–62.
- Guerrini, R., Dobyns, W.B., 2014. Malformations of cortical development: clinical features and genetic causes. *Lancet Neurol* 13:710–726.
- Guo, J., Tan, J., Yang, Y., Zhou, H., Hu, S., Hashan, A., et al., 2014. Variation and signatures of selection on the human face. *J Hum Evol* 75:143–152.
- Guyot, L., Dubuc, M., Richard, O., Philip, N., Dutour, O., 2003. Comparison between direct clinical and digital photogrammetric measurements in patients with 22q11 microdeletion. *Int J Oral Maxillofac Surg* 32:246–252.
- Gwilliam, J.R., Cunningham, S.J., Hutton, T., 2006. Reproducibility of soft tissue landmarks on three-dimensional facial scans. *Eur J Orthod* 28:408–415.
- Hall, B.D., Graham, J.M., Cassidy, S.B., Opitz, J.M., 2009. Elements of morphology: Standard terminology for the periorbital region. *Am J Med Genet A* 149A:29–39.

- Hall, J.G., Allanson, J.E., Gripp, K.W., Slavotinek, A.M., 2007. Handbook of physical measurements. Oxford University Press, Oxford; New York.
- Hallgrímsson, B., Percival, C.J., Green, R., Young, N.M., Mio, W., Marcucio, R., 2015. Morphometrics, 3D Imaging, and Craniofacial Development. *Curr Top Dev Biol* 115:561–597.
- Hammond, P., 2007. The use of 3D face shape modelling in dysmorphology. *Arch Dis Child* 92:1120–1126.
- Hammond, P., Forster-Gibson, C., Chudley, A.E., Allanson, J.E., Hutton, T.J., Farrell, S.A., et al., 2008. Face-brain asymmetry in autism spectrum disorders. *Mol Psychiatry* 13:614–623.
- Hammond, P., Forster-Gibson, C., Chudley, A.E., Allanson, J.E., Hutton, T.J., Farrell, S.A., et al., 2008. Face-brain asymmetry in autism spectrum disorders. *Mol Psychiatry* 13:614–623.
- Hammond, P., Hannes, F., Suttie, M., Devriendt, K., Vermeesch, J.R., Faravelli, F., et al., 2012a. Fine-grained facial phenotype-genotype analysis in Wolf-Hirschhorn syndrome. *Eur J Hum Genet* 20:33–40.
- Hammond, P., Hutton, T.J., Allanson, J.E., Buxton, B., Campbell, L.E., Clayton-Smith, J., et al., 2005. Discriminating power of localized three-dimensional facial morphology. *Am J Hum Genet* 77:999–1010.
- Hammond, P., Hutton, T.J., Allanson, J.E., Campbell, L.E., Hennekam, R.C.M., Holden, S., et al., 2004. 3D analysis of facial morphology. *Am J Med Genet A* 126A:339–348.
- Hammond, P., McKee, S., Suttie, M., Allanson, J., Cobben, J.-M., Maas, S.M., et al., 2014. Opposite effects on facial morphology due to gene dosage sensitivity. *Hum Genet* 133:1117–1125.
- Hammond, P., Suttie, M., 2012. Large-scale objective phenotyping of 3D facial morphology. *Human Mutation*. 33:817-25.
- Hammond, P., Suttie, M., Hennekam, R.C., Allanson, J., Shore, E.M., Kaplan, F.S., 2012b. The face signature of fibrodysplasia ossificans progressiva. *Am J Med Genet A* 158A:1368–1380.
- Hart, T.C., Hart, P.S., 2009. Genetic studies of craniofacial anomalies: clinical implications and applications. *Orthod Craniofac Res* 12:212–220.
- Haumaitre, C., Barbacci, E., Jenny, M., Ott, M.O., Gradwohl, G., Cereghini, S., 2005. Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. *Proc Natl Acad Sci USA* 102:1490–1495.
- Heike, C.L., Cunningham, M.L., Hing, A.V., Stuhaug, E., Starr, J.R., 2009. Picture perfect? Reliability of craniofacial anthropometry using three-dimensional digital stereophotogrammetry. *Plast Reconstr Surg* 124:1261–1272.
- Heike, C.L., Upson, K., Stuhaug, E., Weinberg, S.M., 2010. 3D digital stereophotogrammetry: a practical guide to facial image acquisition. *Head Face Med* 6:18.

- Heinzen, E.L., Radtke, R.A., Urban, T.J., Cavalleri, G.L., Depondt, C., Need, A.C., et al., 2010. Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *Am J Hum Genet* 86:707–718.
- Helbig, I., Scheffer, I.E., Mulley, J.C., Berkovic, S.F., 2008. Navigating the channels and beyond: unravelling the genetics of the epilepsies. *Lancet Neurol* 7:231–245.
- Helbig, I., Swinkels, M.E.M., Aten, E., Caliebe, A., van't Slot, R., Boor, R., et al., 2014. Structural genomic variation in childhood epilepsies with complex phenotypes. *Eur. J. Hum Genet* 22:896–901.
- Hennekam, R.C., Krantz, I.D., Allanson, J.E., 2010. *Gorlin's syndromes of the head and neck*, 5th ed. ed. Oxford University Press, Oxford; New York.
- Hennekam, R.C.M., Cormier-Daire, V., Hall, J.G., Méhes, K., Patton, M., Stevenson, R.E., 2009. Elements of morphology: Standard terminology for the nose and philtrum. *Am J Med Genet A* 149A:61–76.
- Hennessy, R.J., Baldwin, P.A., Browne, D.J., Kinsella, A., Waddington, J.L., 2010. Frontonasal dysmorphology in bipolar disorder by 3D laser surface imaging and geometric morphometrics: comparisons with schizophrenia. *Schizophr Res* 122:63–71.
- Hennessy, R.J., Baldwin, P.A., Browne, D.J., Kinsella, A., Waddington, J.L., 2007. Three-dimensional laser surface imaging and geometric morphometrics resolve frontonasal dysmorphology in schizophrenia. *Biol Psychiatry* 61:1187–1194.
- Hennessy, R.J., Kinsella, A., Waddington, J.L., 2002. 3D laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebro-craniofacial morphogenesis: initial application to sexual dimorphism. *Biol Psychiatry* 51:507–514.
- Hennessy, R.J., Lane, A., Kinsella, A., Larkin, C., O'Callaghan, E., Waddington, J.L., 2004. 3D morphometrics of craniofacial dysmorphology reveals sex-specific asymmetries in schizophrenia. *Schizophr Res* 67:261–268.
- Hennessy, R.J., Stringer, C.B., 2002. Geometric morphometric study of the regional variation of modern human craniofacial form. *Am J Phys Anthropol* 117:37–48.
- Heulens, I., Suttie, M., Postnov, A., De Clerck, N., Perrotta, C.S., Mattina, T., et al., 2013. Craniofacial characteristics of fragile X syndrome in mouse and man. *Eur J Hum Genet* 21:816–823.
- Hiesberger, T., Bai, Y., Shao, X., McNally, B.T., Sinclair, A.M., Tian, X., et al., 2004. Mutation of hepatocyte nuclear factor-1beta inhibits Pkhd1 gene expression and produces renal cysts in mice. *J Clin Invest* 113:814–825.
- Honrado, C.P., Larrabee, W.F., 2004. Update in three-dimensional imaging in facial plastic surgery. *Curr Opin Otolaryngol Head Neck Surg* 12:327–331.
- Hopman, S.M.J., Merks, J.H.M., Suttie, M., Hennekam, R.C.M., Hammond, P., 2014. Face shape differs in phylogenetically related populations. *Eur J Hum Genet* 22:1268–1271.
- Horikawa, Y., Iwasaki, N., Hara, M., Furuta, H., Hinokio, Y., Cockburn, B.N., et al., 1997. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet* 17:384–385.

- Huang, C.S., Liu, X.Q., Chen, Y.R., 2013. Facial asymmetry index in normal young adults. *Orthod Craniofac Res* 16:97–104.
- Hunter, A., Frias, J.L., Gillessen-Kaesbach, G., Hughes, H., Jones, K.L., Wilson, L., 2009. Elements of morphology: Standard terminology for the ear. *Am J Med Genet A* 149A:40–60.
- Hunter, A.G., 1996. Craniofacial anthropometric analysis in several types of chondrodysplasia. *Am J Med Genet* 65:5–12.
- Hutton, T., 2004. Dense surface models of the human face. UCL, London.
- Hutton, T.J., Buxton, B.F., Hammond, P., Potts, H.W.W., 2003. Estimating average growth trajectories in shape-space using kernel smoothing. *IEEE Trans Med Imaging* 22:747–753.
- Hutton, T.J., Buxton, B.R., Hammond, P., 2001. Dense surface point distribution models of the human face, in: *Mathematical Methods in Biomedical Image Analysis*. IEEE Comput Soc, pp. 153–160.
- Iafate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., et al., 2004. Detection of large-scale variation in the human genome. *Nat Genet* 36:949–951.
- International League Against Epilepsy Consortium on Complex Epilepsies, 2014. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol* 13:893–903.
- Israel, H., 1973. Recent knowledge concerning craniofacial aging. *Angle Orthod* 43:176–184.
- Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., et al., 2009. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 84:148–161.
- Jamison, P.L., Ward, R.E., 1993. Brief communication: measurement size, precision, and reliability in craniofacial anthropometry: bigger is better. *Am J Phys Anthropol* 90:495–500.
- Janssen, I., Heymsfield, S.B., Allison, D.B., Kotler, D.P., Ross, R., 2002. Body mass index and waist circumference independently contribute to the prediction of nonabdominal, abdominal subcutaneous, and visceral fat. *Am J Clin Nutr* 75:683–688.
- Johannsdottir, B., Thorarinsson, F., Thordarson, A., Magnusson, T.E., 2005. Heritability of craniofacial characteristics between parents and offspring estimated from lateral cephalograms. *Am J Orthod Dentofacial Orthop* 127:200–207–261.
- Johnston, D.J., Millett, D.T., Ayoub, A.F., Bock, M., 2003. Are facial expressions reproducible? *Cleft Palate Craniofac J* 40:291–296.
- Jones, K.L., 2006. *Smith's Recognizable Patterns of Human Malformation*, 6th ed. Saunders.
- Kalachikov, S., Evgrafov, O., Ross, B., Winawer, M., Barker-Cummings, C., Martinelli Boneschi, F., et al., 2002. Mutations in *LGI1* cause autosomal-dominant partial epilepsy with auditory features. *Nat Genet* 30:335–341.

- Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., et al., 2011. Spatio-temporal transcriptome of the human brain. *Nature* 478:483–489.
- Kasperavičiūtė, D., Catarino, C.B., Chinthapalli, K., Clayton, L.M.S., Thom, M., Martinian, L., et al., 2011. Uncovering genomic causes of co-morbidity in epilepsy: gene-driven phenotypic characterization of rare microdeletions. *PLoS One* 6:e23182.
- Kasperaviciute, D., Catarino, C.B., Matarin, M., Leu, C., Novy, J., Tostevin, A., et al., 2013. Epilepsy, hippocampal sclerosis and febrile seizures linked by common genetic variation around SCN1A. *Brain* 136:3140-50.
- Kass, H.R., Winesett, S.P., Bessone, S.K., Turner, Z., Kossoff, E.H., 2016. Use of dietary therapies amongst patients with GLUT1 deficiency syndrome. *Seizure* 35:83–87.
- Katina, S., McNeil, K., Ayoub, A., Guilfoyle, B., Khambay, B., Siebert, P., et al., 2016. The definitions of three-dimensional landmarks on the human face: an interdisciplinary view. *J Anat* 228:355–365.
- Kau, C.H., Richmond, S., Incrapera, A., English, J., Xia, J.J., 2007. Three-dimensional surface acquisition systems for the study of facial morphology and their application to maxillofacial surgery. *Int J Med Robot* 3:97–110.
- Kau, C.H., Richmond, S., Zhurov, A., Ovsenik, M., Tawfik, W., Borbely, P., et al., 2010. Use of 3-dimensional surface acquisition to study facial morphology in 5 populations. *Am J Orthod Dentofacial Orthop* 137:S56.e1-9.
- Kau, C.H., Richmond, S., Zhurov, A.I., Knox, J., Chestnutt, I., Hartles, F., et al., 2005. Reliability of measuring facial morphology with a 3-dimensional laser scanning system. *Am J Orthod Dentofacial Orthop* 128:424–430.
- Kawai, T., Natsume, N., Shibata, H., Yamamoto, T., 1990a. Three-dimensional analysis of facial morphology using moiré stripes. Part I. Method. *Int J Oral Maxillofac Surg* 19:356–358.
- Kawai, T., Natsume, N., Shibata, H., Yamamoto, T., 1990b. Three-dimensional analysis of facial morphology using moiré stripes. Part II. Analysis of normal adults. *Int J Oral Maxillofac Surg* 19:359–362.
- Keys, A., Fidanza, F., Karvonen, M.J., Kimura, N., Taylor, H.L., 1972. Indices of relative weight and obesity. *J Chronic Dis* 25:329–343.
- Khambay, B., Nairn, N., Bell, A., Miller, J., Bowman, A., Ayoub, A.F., 2008. Validation and reproducibility of a high-resolution three-dimensional facial imaging system. *Br J Oral Maxillofac Surg* 46:27–32.
- Kharbanda, M., Tolmie, J., Joss, S., 2015. How to use... microarray comparative genomic hybridisation to investigate developmental disorders. *Arch Dis Child Educ Pract Ed* 100:24–29.
- Knoops, P.G., Beaumont, C.A., Borghi, A., Rodriguez-Florez, N., Breakey, R.W., Rodgers, W., et al., 2017. Comparison of three-dimensional scanner systems for craniomaxillofacial imaging. *J Plast Reconstr Aesthet Surg* 70:441-449.
- Kohn, L.A.P., 1991. The Role of Genetics in Craniofacial Morphology and Growth. *Annu Rev Anthropol* 20:261–278.

- Kook, M., Jung, S., Park, H., Oh, H., Ryu, S., Cho, J., et al., 2014. A comparison study of different facial soft tissue analysis methods. *J Craniomaxillofac Surg* 42:648–656.
- Kurahashi, H., Wang, J., Ishii, A., Kojima, T., Wakai, S., Kizawa, T., et al., 2009. Deletions involving both KCNQ2 and CHRNA4 present with benign familial neonatal seizures. *Neurology* 73:1214–1217.
- Lal, D., Ruppert, A., Trucks, H., Schulz, H., de Kovel, C.G., Kasteleijn-Nolst Trenité, D., et al., 2015. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. *PLoS Genet.* 11:e1005226.
- Lam, B., Ip, M.S.M., Tench, E., Ryan, C.F., 2005. Craniofacial profile in Asian and white subjects with obstructive sleep apnoea. *Thorax* 60:504–510.
- Lane, A., Kinsella, A., Murphy, P., Byrne, M., Keenan, J., Colgan, K., et al., 1997. The anthropometric assessment of dysmorphic features in schizophrenia as an index of its developmental origins. *Psychol Med* 27:1155–1164.
- Lane, C., Harrell, W., 2008. Completing the 3-dimensional picture. *Am J Orthod Dentofacial Orthop* 133:612–620.
- Lawn, N.D., Bamlet, W.R., Radhakrishnan, K., O'Brien, P.C., So, E.L., 2004. Injuries due to seizures in persons with epilepsy: a population-based study. *Neurology* 63:1565–1570.
- Le Douarin, N.M., Couly, G., Creuzet, S.E., 2012. The neural crest is a powerful regulator of pre-otic brain development. *Dev Biol* 366:74–82.
- Lee, J.Y., Han, Q., Trotman, C.-A., 2004. Three-dimensional facial imaging: accuracy and considerations for clinical applications in orthodontics. *Angle Orthod* 74:587–593.
- Lee, R.W.W., Vasudavan, S., Hui, D.S., Prvan, T., Petocz, P., Darendeliler, M.A., et al., 2010. Differences in craniofacial structures and obesity in Caucasian and Chinese patients with obstructive sleep apnea. *Sleep* 33:1075–1080.
- Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., et al., 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–176.
- Lele, S., Richtsmeier, J.T., 1995. Euclidean distance matrix analysis: confidence intervals for form and growth differences. *Am J Phys Anthropol* 98:73–86.
- Lemke, J.R., Riesch, E., Scheurenbrand, T., Schubach, M., Wilhelm, C., Steiner, I., et al., 2012. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia* 53:1387–1398.
- Leopold, D.A., Rhodes, G., 2010. A comparative view of face perception. *J Comp Psychol* 124:233–251.
- Leu, C., Coppola, A., Sisodiya, S.M., 2016. Progress from genome-wide association studies and copy number variant studies in epilepsy. *Curr Opin Neurol* 29:158–167.
- Lin, J., Chiou, W., Weng, H., Tsai, Y., Liu, T., 2002. Comparison of three-dimensional anthropometric body surface scanning to waist-hip ratio and body mass index in correlation with metabolic risk factors. *J Clin Epidemiol* 55:757–766.

- Lincoln, K.P., Sun, A.Y.T., Prihoda, T.J., Sutton, A.J., 2015. Comparative Accuracy of Facial Models Fabricated Using Traditional and 3D Imaging Techniques. *J Prosthodont* 25:207-15.
- Lindell, A.K., Hudry, K., 2013. Atypicalities in cortical structure, handedness, and functional lateralization for language in autism spectrum disorders. *Neuropsychol Rev* 23:257-270.
- Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M.M., Hysi, P.G., et al., 2012. A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet* 8:e1002932.
- Lowther, C., Costain, G., Stavropoulos, D.J., Melvin, R., Silversides, C.K., Andrade, D.M., et al., 2015. Delineating the 15q13.3 microdeletion phenotype: a case series and comprehensive review of the literature. *Genet Med* 17:149-157.
- Lübbers, H., Medinger, L., Kruse, A., Grätz, K.W., Matthews, F., 2010. Precision and accuracy of the 3dMD photogrammetric system in craniomaxillofacial application. *J Craniofac Surg* 21:763-767.
- Lübbers, H., Medinger, L., Kruse, A.L., Grätz, K.W., Obwegeser, J.A., Matthews, F., 2012. The influence of involuntary facial movements on craniofacial anthropometry: a survey using a three-dimensional photographic system. *Br J Oral Maxillofac Surg* 50:171-175.
- Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., et al., 1991. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* 66:219-232.
- Ma, L., Xu, T., Lin, J., 2009. Validation of a three-dimensional facial scanning system based on structured light techniques. *Comput Methods Programs Biomed* 94:290-298.
- Maal, T.J.J., van Loon, B., Plooi, J.M., Rangel, F., Ettema, A.M., Borstlap, W.A., et al., 2010. Registration of 3-dimensional facial photographs for clinical use. *J Oral Maxillofac Surg* 68:2391-2401.
- Maal, T.J.J., Verhamme, L.M., van Loon, B., Plooi, J.M., Rangel, F.A., Kho, A., et al., 2011. Variation of the face in rest using 3D stereophotogrammetry. *Int J Oral Maxillofac Surg* 40:1252-1257.
- Machado, L.R., Ottolini, B., 2015. An evolutionary history of defensins: a role for copy number variation in maximizing host innate and adaptive immune responses. *Front Immunol* 6:115.
- Madden, M.C., Karlan, M.S., 1979. Moiré photography as a means of topographical mapping of the human face. *Ann Biomed Eng* 7:95-102.
- Maeda, T., Ohno, M., Matsunobu, A., Yoshihara, K., Yabe, N., 1991. A cytogenetic survey of 14,835 consecutive liveborns. *Jinrui Idengaku Zasshi* 36:117-129.
- Maguire, M., 2009. The birth of biometric security. *Anthropology Today* 25:9-14.
- Mainous, A.G., Johnson, S.P., Chirina, S., Baker, R., 2013. Academic family physicians' perception of genetic testing and integration into practice: a CERA study. *Fam Med* 45:257-262.

- Makki, N., Capecchi, M.R., 2011. Identification of novel Hoxa1 downstream targets regulating hindbrain, neural crest and inner ear development. *Dev Biol* 357:295–304.
- Malmgren, K., Thom, M., 2012. Hippocampal sclerosis--origins and imaging. *Epilepsia* 53 Suppl 4:19–33.
- Marcucio, R., Hallgrímsson, B., Young, N.M., 2015. Facial Morphogenesis: Physical and Molecular Interactions Between the Brain and the Face. *Curr Top Dev Biol* 115:299–320.
- Marcucio, R.S., Cordero, D.R., Hu, D., Helms, J.A., 2005. Molecular interactions coordinating the development of the forebrain and face. *Dev Biol* 284:48–61.
- Marcucio, R.S., Young, N.M., Hu, D., Hallgrímsson, B., 2011. Mechanisms that underlie co-variation of the brain and face. *Genesis* 49:177–189.
- Marečková, K., Chakravarty, M.M., Lawrence, C., Leonard, G., Perusse, D., Perron, M., et al., 2015. Identifying craniofacial features associated with prenatal exposure to androgens and testing their relationship with brain development. *Brain Struct Funct* 220:3233–3244.
- Martínez-Abadías, N., Esparza, M., Sjøvold, T., González-José, R., Santos, M., Hernández, M., 2009. Heritability of human cranial dimensions: comparing the evolvability of different cranial regions. *J Anat* 214:19–35.
- Maulisova, A., Korman, B., Rey, G., Bernal, B., Duchowny, M., Niederlova, M., et al., 2016. Atypical language representation in children with intractable temporal lobe epilepsy. *Epilepsy Behav* 58:91–96.
- Maves, S.N., Williams, M.S., Williams, J.L., Levonian, P.J., Josephson, K.D., 2007. Analysis of 88 adult patients referred for genetics evaluation. *Am J Med Genet C Semin Med Genet* 145C:232–240.
- Mefford, H.C., Clauin, S., Sharp, A.J., Moller, R.S., Ullmann, R., Kapur, R., et al., 2007. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet* 81:1057–1069.
- Mefford, H.C., Muhle, H., Ostertag, P., von Spiczak, S., Buysse, K., Baker, C., et al., 2010. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet*. 6:e1000962.
- Merlob, P., Sivan, Y., Reisner, S.H., 1984. Anthropometric measurements of the newborn infant (27 to 41 gestational weeks). *Birth Defects Orig. Artic. Ser.* 20:1–52.
- Metzger, T.E., Kula, K.S., Eckert, G.J., Ghoneima, A.A., 2013. Orthodontic soft-tissue parameters: a comparison of cone-beam computed tomography and the 3dMD imaging system. *Am J Orthod Dentofacial Orthop* 144:672–681.
- Metzler, P., Geiger, E.J., Chang, C.C., Steinbacher, D.M., 2014. Surgically assisted maxillary expansion imparts three-dimensional nasal change. *J Oral Maxillofac Surg* 72:2005–2014.
- Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., et al., 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic

test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum Genet* 86:749–764.

Ming, J.E., Muenke, M., 1998. Holoprosencephaly: from Homer to Hedgehog. *Clin Genet* 53:155–163.

Møller, R.S., Heron, S.E., Larsen, L.H.G., Lim, C.X., Ricos, M.G., Bayly, M.A., et al., 2015. Mutations in *KCNT1* cause a spectrum of focal epilepsies. *Epilepsia* 56:e114-120.

Monaco, F., Mula, M., 2011. Cesare Lombroso and epilepsy 100 years later: an unabridged report of his original transactions. *Epilepsia* 52:679–688.

Moreno De-Luca, D., Mulle, J.G., Kaminsky, E.B., Sanders, S.J., Myers, S.M., Adam, M.P., et al., 2010. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet* 87:618–630.

Mullen, S.A., Carvill, G.L., Bellows, S., Bayly, M.A., Trucks, H., Lal, D., et al., 2013. Copy number variants are frequent in genetic generalized epilepsy with intellectual disability. *Neurology* 81:1507–1514.

Mykytyn, K., Nishimura, D.Y., Searby, C.C., Shastri, M., Yen, H., Beck, J.S., et al., 2002. Identification of the gene (*BBS1*) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat Genet* 31:435–438.

Nagamani, S.C.S., Erez, A., Shen, J., Li, C., Roeder, E., Cox, S., et al., 2010. Clinical spectrum associated with recurrent genomic rearrangements in chromosome 17q12. *Eur J Hum Genet* 18:278–284.

Nakajima, J., Okamoto, N., Tohyama, J., Kato, M., Arai, H., Funahashi, O., et al., 2015. De novo *EEF1A2* mutations in patients with characteristic facial features, intellectual disability, autistic behaviors and epilepsy. *Clin Genet* 87:356–361.

Neligan, A., Bell, G.S., Johnson, A.L., Goodridge, D.M., Shorvon, S.D., Sander, J.W., 2011. The long-term risk of premature mortality in people with epilepsy. *Brain* 134:388–395.

Ngugi, A.K., Bottomley, C., Kleinschmidt, I., Sander, J.W., Newton, C.R., 2010. Estimation of the burden of active and life-time epilepsy: a meta-analytic approach. *Epilepsia* 51:883–890.

Ngugi, A.K., Kariuki, S.M., Bottomley, C., Kleinschmidt, I., Sander, J.W., Newton, C.R., 2011. Incidence of epilepsy: a systematic review and meta-analysis. *Neurology* 77:1005–1012.

Nielsen, J., Wohlert, M., 1991. Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. *Hum Genet* 87:81–83.

Novy, J., Catarino, C.B., Chinthapalli, K., Smith, S.M., Clayton-Smith, J., Hennekam, R.C.M., et al., 2012. Another cause of vaccine encephalopathy: a case of Angelman syndrome. *Eur J Med Genet* 55:338–341.

O'Boyle, K.H., Gallagher, F.D., O'Sullivan, M., McDevitt, W.E., 1996. The effect of posture change on the position of the skin marks for the transverse horizontal axis. *J Prosthet Dent* 75:545–551.

Olson, H., Shen, Y., Avallone, J., Sheidley, B.R., Pinsky, R., Bergin, A.M., et al., 2014. Copy number variation plays an important role in clinical epilepsy. *Ann Neurol* 75:943–958.

Oram, R.A., Edghill, E.L., Blackman, J., Taylor, M.J.O., Kay, T., Flanagan, S.E., et al., 2010. Mutations in the hepatocyte nuclear factor-1 $\beta$  (HNF1B) gene are common with combined uterine and renal malformations but are not found with isolated uterine malformations. *Am J Obstet Gynecol* 203:364.e1-5.

Orup, H.I., Holmes, L.B., Keith, D.A., Coull, B.A., 2003. Craniofacial skeletal deviations following in utero exposure to the anticonvulsant phenytoin: monotherapy and polytherapy. *Orthod Craniofac Res* 6:2–19.

Othman, S.A., Ahmad, R., Mericant, A.F., Jamaludin, M., 2013. Reproducibility of facial soft tissue landmarks on facial images captured on a 3D camera. *Aust Orthod J* 29:58–65.

Pandolfo, M., 2013. Pediatric epilepsy genetics. *Curr Opin Neurol* 26:137–145.

Parssinen, T.M., 1974. Popular science and society: The phrenology movement in early Victorian Britain. *Journal of Social History* 8:1–20.

Paternoster, L., Zhurov, A.I., Toma, A.M., Kemp, J.P., St Pourcain, B., Timpson, N.J., et al., 2012. Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am J Hum Genet* 90:478–485.

Pecora, N.G., Baccetti, T., McNamara, J.A., 2008. The aging craniofacial complex: a longitudinal cephalometric study from late adolescence to late adulthood. *Am J Orthod Dentofacial Orthop* 134:496–505.

Pelin, C., Zağyapan, R., Yazici, C., Kürkçüoğlu, A., 2010. Body height estimation from head and face dimensions: a different method. *J Forensic Sci* 55:1326–1330.

Perucca, E., 2005. An introduction to antiepileptic drugs. *Epilepsia* 46 Suppl 4:31–37.

Plooi, J.M., Swennen, G.R.J., Rangel, F.A., Maal, T.J.J., Schutyser, F.C., Bronkhorst, E.M., et al., 2009. Evaluation of reproducibility and reliability of 3D soft tissue analysis using 3D stereophotogrammetry. *Int J Oral Maxillofac Surg* 38:267–273.

Pound, N., Lawson, D.W., Toma, A.M., Richmond, S., Zhurov, A.I., Penton-Voak, I.S., 2014. Facial fluctuating asymmetry is not associated with childhood ill-health in a large British cohort study. *Proc Biol Sci* 281:1792.

Prasad, S., Katina, S., Hennessy, R.J., Murphy, K.C., Bowman, A.W., Waddington, J.L., 2015. Craniofacial dysmorphology in 22q11.2 deletion syndrome by 3D laser surface imaging and geometric morphometrics: illuminating the developmental relationship to risk for psychosis. *Am J Med Genet A* 167A:529–536.

Quan, M., Fadl, A., Small, K., Tepper, O., Kumar, N., Choi, M., et al., 2011. Defining pseudoptosis (bottoming out) 3 years after short-scar medial pedicle breast reduction. *Aesthetic Plast Surg* 35:357–364.

Ras, F., Habets, L.L., van Ginkel, F.C., Prahl-Andersen, B., 1996. Quantification of facial morphology using stereophotogrammetry--demonstration of a new concept. *J Dent* 24:369–374.

- Ras, F., Habets, L.L., van Ginkel, F.C., Prahl-Andersen, B., 1995. Method for quantifying facial asymmetry in three dimensions using stereophotogrammetry. *Angle Orthod* 65:233–239.
- Raymond, A.A., Fish, D.R., Sisodiya, S.M., Alsanjari, N., Stevens, J.M., Shorvon, S.D., 1995. Abnormalities of gyration, heterotopias, tuberous sclerosis, focal cortical dysplasia, microdysgenesis, dysembryoplastic neuroepithelial tumour and dysgenesis of the archicortex in epilepsy. Clinical, EEG and neuroimaging features in 100 adult patients. *Brain* 118:629–660.
- Reardon, W., Donnai, D., 2007. Dysmorphology demystified. *Arch. Dis. Child. Fetal Neonatal Ed* 92:F225-229.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., et al., 2006. Global variation in copy number in the human genome. *Nature* 444:444–454.
- Rhodes, G., 2006. The evolutionary psychology of facial beauty. *Annu Rev Psychol* 57:199–226.
- Richtsmeier, J.T., DeLeon, V.B., Lele, S.R., 2002. The promise of geometric morphometrics. *Am J Phys Anthropol Suppl* 35:63–91.
- Richtsmeier, J.T., Flaherty, K., 2013. Hand in glove: brain and skull in development and dysmorphogenesis. *Acta Neuropathol* 125:469–489.
- Riphagen, J.M., van Neck, J.W., van Adrichem, L.N.A., 2008. 3D surface imaging in medicine: a review of working principles and implications for imaging the unsedated child. *J Craniofac Surg* 19:517–524.
- Rivera, B., Gayden, T., Carrot-Zhang, J., Nadaf, J., Boshari, T., Faury, D., et al., 2016. Germline and somatic FGFR1 abnormalities in dysembryoplastic neuroepithelial tumors. *Acta Neuropathol* 131:847–863.
- Robin, N.H., 2011. Genetic drift. The mall test (or fun with a dysmorphologist). *Am J Med Genet A* 155A:2909.
- Robinow, M., Roche, A.F., 1973. Low-set ears. *Am. J. Dis. Child.* 125:482–483.
- Roelandt, P., Antoniou, A., Libbrecht, L., Van Steenberghe, W., Laleman, W., Verslype, C., et al., 2012. HNF1B deficiency causes ciliary defects in human cholangiocytes. *Hepatology* 56:1178–1181.
- Roostaeian, J., Adams, W.P., 2014. Three-Dimensional Imaging for Breast Augmentation: Is This Technology Providing Accurate Simulations? *Aesthet Surg J* 34:857–875.
- Rosati, R., De Menezes, M., da Silva, A.M.B.R., Rossetti, A., Lanza Attisano, G.C., Sforza, C., 2014. Stereophotogrammetric evaluation of tooth-induced labial protrusion. *J Prosthodont* 23:347–352.
- Rosati, R., De Menezes, M., Rossetti, A., Sforza, C., Ferrario, V.F., 2010. Digital dental cast placement in 3-dimensional, full-face reconstruction: a technical evaluation. *Am J Orthod Dentofacial Orthop* 138:84–88.

- Rosati, R., Rossetti, A., De Menezes, M., Ferrario, V.F., Sforza, C., 2012. The occlusal plane in the facial context: inter-operator repeatability of a new three-dimensional method. *Int J Oral Sci* 4:34–37.
- Rossetti, A., De Menezes, M., Rosati, R., Ferrario, V.F., Sforza, C., 2013. The role of the golden proportion in the evaluation of facial esthetics. *Angle Orthod* 83:801–808.
- Salm, M., Abbate, K., Appelbaum, P., Ottman, R., Chung, W., Marder, K., et al., 2014. Use of genetic tests among neurologists and psychiatrists: knowledge, attitudes, behaviors, and needs for training. *J Genet Couns* 23:156–163.
- Salvador-Carulla, L., Reed, G.M., Vaez-Azizi, L.M., Cooper, S.-A., Martinez-Leal, R., Bertelli, M., et al., 2011. Intellectual developmental disorders: towards a new name, definition and framework for “mental retardation/intellectual disability” in ICD-11. *World Psychiatry* 10:175–180.
- Sant’Anna, L.B., Tosello, D.O., 2006. Fetal alcohol syndrome and developing craniofacial and dental structures--a review. *Orthod Craniofac Res* 9:172–185.
- Savoye, I., Loos, R., Carels, C., Derom, C., Vlietinck, R., 1998. A genetic study of anteroposterior and vertical facial proportions using model-fitting. *Angle Orthod* 68:467–470.
- Sawyer, A.R., See, M., Nduka, C., 2009. Assessment of the reproducibility of facial expressions with 3-D stereophotogrammetry. *Otolaryngol Head Neck Surg* 140:76–81.
- Schaaf, H., Pons-Kuehnemann, J., Malik, C.Y., Streckbein, P., Preuss, M., Howaldt, H.-P., et al., 2010. Accuracy of three-dimensional photogrammetric images in non-synostotic cranial deformities. *Neuropediatrics* 41:24–29.
- Scheffer, I.E., Berkovic, S., Capovilla, G., Connolly, M.B., French, J., Guilhoto, L., et al., 2017. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* 58:512–521.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., et al., 2004. Large-scale copy number polymorphism in the human genome. *Science* 305:525–528.
- See, M.S., Roberts, C., Nduka, C., 2008. Age- and gravity-related changes in facial morphology: 3-dimensional analysis of facial morphology in mother-daughter pairs. *J Oral Maxillofac Surg* 66:1410–1416.
- Sforza, C., Grandi, G., Binelli, M., Dolci, C., De Menezes, M., Ferrario, V.F., 2010. Age- and sex-related changes in three-dimensional lip morphology. *Forensic Sci Int* 200:182.e1-7.
- Sforza, C., Grandi, G., De Menezes, M., Tartaglia, G.M., Ferrario, V.F., 2011. Age- and sex-related changes in the normal human external nose. *Forensic Sci Int* 204:205.e1-9.
- Shaffer, L.G., Lupski, J.R., 2000. Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu Rev Genet* 34:297–329.
- Shaner, D.J., Peterson, A.E., Beattie, O.B., Bamforth, J.S., 2000. Assessment of soft tissue facial asymmetry in medically normal and syndrome-affected individuals by analysis of landmarks and measurements. *Am J Med Genet* 93:143–154.

- Shapira, S.K., 1998. An update on chromosome deletion and microdeletion syndromes. *Curr Opin Pediatr* 10:622–627.
- Sharp, A.J., Cheng, Z., Eichler, E.E., 2006. Structural variation of the human genome. *Annu Rev Genomics Hum Genet* 7:407–442.
- Sillanpää, M., Jalava, M., Kaleva, O., Shinnar, S., 1998. Long-term prognosis of seizures with onset in childhood. *N Engl J Med* 338:1715–1722.
- Simpson, D., 2005. Phrenology and the neurosciences: contributions of F. J. Gall and J. G. Spurzheim. *ANZ J Surg* 75:475–482.
- Singh, R., Gardner, R.J.M., Crossland, K.M., Scheffer, I.E., Berkovic, S.F., 2002. Chromosomal abnormalities and epilepsy: a review for clinicians and gene hunters. *Epilepsia* 43:127–140.
- Sinha, P., Balas, B., Ostrovsky, Y., Russell, R., 2006. Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE* 94:1948–1962.
- Sisodiya, S., 2008. Brain structure, function, and genetics revealed by studies of the eye and face. *Curr Opin Neurol* 21:404–409.
- Sisodiya, S.M., 2004. Malformations of cortical development: burdens and insights from important causes of human epilepsy. *Lancet Neurol* 3:29–38.
- Sisodiya, S.M., Mefford, H.C., 2011. Genetic contribution to common epilepsies. *Curr Opin Neurol* 24:140–145.
- Smeets, D., Claes, P., Vandermeulen, D., Clement, J.G., 2010. Objective 3D face recognition: Evolution, approaches and challenges. *Forensic Sci Int* 201:125–132.
- Smith, D.W., 1966. Dysmorphology (teratology). *J Pediatr* 69:1150–1169.
- Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., et al., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17:479–489.
- Som, P.M., Naidich, T.P., 2013. Illustrated review of the embryology and development of the facial region, part 1: Early face and lateral nasal cavities. *AJNR Am J Neuroradiol* 34:2233–2240.
- Stemmler, J., 1993. The physiognomical portraits of Johann Caspar Lavater. *The Art Bulletin* 75:151–168.
- Striano, P., Coppola, A., Paravidino, R., Malacarne, M., Gimelli, S., Robbiano, A., et al., 2012. Clinical significance of rare copy number variations in epilepsy: a case-control survey using microarray-based comparative genomic hybridization. *Arch Neurol* 69:322–330.
- Sukno, F.M., Waddington, J.L., Whelan, P.F., 2015. 3-D Facial Landmark Localization With Asymmetry Patterns and Shape Regression from Incomplete Local Features. *IEEE Transactions on Cybernetics* 45:1717–1730.

Suttie, M., Foroud, T., Wetherill, L., Jacobson, J.L., Moltano, C.D., Meintjes, E.M., et al., 2013. Facial dysmorphism across the fetal alcohol spectrum. *Pediatrics* 131:e779-788.

Tassabehji, M., Hammond, P., Karmiloff-Smith, A., Thompson, P., Thorgeirsson, S.S., Durkin, M.E., et al., 2005. GTF2IRD1 in craniofacial development of humans and mice. *Science* 310:1184-1187.

Taylor, M.R.G., Edwards, J.G., Ku, L., 2006. Lost in transition: challenges in the expanding field of adult genetics. *Am J Med Genet C Semin Med Genet* 142C:294-303.

The Economist, 2015. Looking for answers. *The Economist*. Available at: <https://www.economist.com/news/science-and-technology/21672078-face-recognition-technology-can-diagnose-developmental-disorders-looking>

Thiesen, G., Gribel, B.F., Freitas, M.P.M., 2015. Facial asymmetry: a current review. *Dental Press J Orthod* 20:110-125.

Thomas, I.T., Hintz, R.J., Frias, J.L., 1989. New methods for quantitative and qualitative facial studies: an overview. *J Craniofac Genet Dev Biol* 9:107-111.

Thornhill, R., Møller, A.P., 1997. Developmental stability, disease and medicine. *Biol Rev Camb Philos Soc* 72:497-548.

Tiamkao, S., Sawanyawisuth, K., Asawavichienjinda, T., Yaudnopakao, P., Arunpongpaisal, S., Phuttharak, W., et al., 2009. Predictive risk factors of seizure-related injury in persons with epilepsy. *J Neurol Sci* 285:59-61.

Tobin, J.L., Di Franco, M., Eichers, E., May-Simera, H., Garcia, M., Yan, J., et al., 2008. Inhibition of neural crest migration underlies craniofacial dysmorphism and Hirschsprung's disease in Bardet-Biedl syndrome. *Proc Natl Acad Sci USA* 105:6714-6719.

Toma, A.M., Zhurov, A., Playle, R., Ong, E., Richmond, S., 2009. Reproducibility of facial soft tissue landmarks on 3D laser-scanned facial images. *Orthod Craniofac Res* 12:33-42.

Trevisol-Bittencourt, P.C., da Silva, V.R., Molinari, M.A., Troiano, A.R., 1999. Phenytoin as the first option in female epileptic patients? *Arq Neuropsiquiatr* 57:784-786.

Tzou, C.-H.J., Artner, N.M., Pona, I., Hold, A., Placheta, E., Kropatsch, W.G., et al., 2014. Comparison of three-dimensional surface-imaging systems. *J Plast Reconstr Aesthet Surg* 67:489-497.

van Bon, B.W.M., Mefford, H.C., Menten, B., Koolen, D.A., Sharp, A.J., Nillesen, W.M., et al., 2009. Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *J Med Genet* 46:511-523.

van den Broek, M., Beghi, E., 2004. Accidents in patients with epilepsy: types, circumstances, and complications: a European cohort study. *Epilepsia* 45:667-672.

Vendrame, M., Maski, K.P., Chatterjee, M., Heshmati, A., Krishnamoorthy, K., Tan, W.-H., et al., 2010. Epilepsy in Prader-Willi syndrome: clinical characteristics and correlation to genotype. *Epilepsy Behav* 19:306-310.

- Vermeesch, J.R., Balikova, I., Schrandt-Stumpel, C., Fryns, J.-P., Devriendt, K., 2011. The causality of de novo copy number variants is overestimated. *Eur J Hum Genet* 19:1112–1113.
- Verrotti, A., D'Egidio, C., Mohn, A., Coppola, G., Chiarelli, F., 2011. Weight gain following treatment with valproic acid: pathogenetic mechanisms and clinical implications. *Obes Rev* 12:e32-43.
- Walker, M.C., 2015. Hippocampal Sclerosis: Causes and Prevention. *Semin Neurol* 35:193–200.
- Walters, R.G., Jacquemont, S., Valsesia, A., de Smith, A.J., Martinet, D., Andersson, J., et al., 2010. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 463:671–675.
- Wang, J., Kurahashi, H., Ishii, A., Kojima, T., Ohfu, M., Inoue, T., et al., 2008. Microchromosomal deletions involving SCN1A and adjacent genes in severe myoclonic epilepsy in infancy. *Epilepsia* 49:1528–1534.
- Warburton, D., 1991. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet* 49:995–1013.
- Ward, R.E., Jamison, P.L., Farkas, L.G., 1998. Craniofacial variability index: a simple measure of normal and abnormal variation in the head and face. *Am J Med Genet* 80:232–240.
- Watson, E.K., Shickle, D., Qureshi, N., Emery, J., Austoker, J., 1999. The “new genetics” and primary care: GPs’ views on their role and their educational needs. *Fam Pract* 16:420–425.
- Weinberg, S.M., Naidoo, S., Govier, D.P., Martin, R.A., Kane, A.A., Marazita, M.L., 2006. Anthropometric precision and accuracy of digital three-dimensional photogrammetry: comparing the Genex and 3dMD imaging systems with one another and with direct anthropometry. *J Craniofac Surg* 17:477–483.
- Weinberg, S.M., Neiswanger, K., Richtsmeier, J.T., Maher, B.S., Mooney, M.P., Siegel, M.I., et al., 2008. Three-dimensional morphometric analysis of craniofacial shape in the unaffected relatives of individuals with nonsyndromic orofacial clefts: a possible marker for genetic susceptibility. *Am J Med Genet A* 146A:409–420.
- Wells, J.C.K., Treleaven, P., Cole, T.J., 2007. BMI compared with 3-dimensional body shape: the UK National Sizing Survey. *Am J Clin Nutr* 85:419–425.
- West, K.S., McNamara, J.A., 1999. Changes in the craniofacial complex from adolescence to midadulthood: a cephalometric study. *Am J Orthod Dentofacial Orthop* 115:521–532.
- Whitlock, G., Lewington, S., Sherliker, P., Clarke, R., Emberson, J., Halsey, J., et al., 2009. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* 373:1083–1096.
- Wieland, J., Kapitein-de Haan, S., Zitman, F.G., 2014. Psychiatric disorders in outpatients with borderline intellectual functioning: comparison with both outpatients from regular mental health care and outpatients with mild intellectual disabilities. *Can J Psychiatry* 59:213–219.

- Williams, M.S., 2007. Adult dysmorphology: perspectives on approach to diagnosis and care. *Am J Med Genet C Semin Med Genet* 145C:227–229.
- Winder, R.J., Darvann, T.A., McKnight, W., Magee, J.D.M., Ramsay-Baggs, P., 2008. Technical validation of the Di3D stereophotogrammetry surface imaging system. *Br J Oral Maxillofac Surg* 46:33–37.
- Winter, R.M., 1996. What's in a face? *Nat Genet* 12:124–129.
- Wirrell, E.C., 2006. Epilepsy-related injuries. *Epilepsia* 47 Suppl 1:79–86.
- Wolyniak, M.J., Bemis, L.T., Prunuske, A.J., 2015. Improving medical students' knowledge of genetic disease: a review of current and emerging pedagogical practices. *Adv Med Educ Pract* 6:597–607.
- Wong, J.Y., Oh, A.K., Ohta, E., Hunt, A.T., Rogers, G.F., Mulliken, J.B., et al., 2008. Validity and reliability of craniofacial anthropometric measurement of 3D digital photogrammetric images. *Cleft Palate Craniofac. J.* 45:232–239.
- Zheng, W., McLerran, D.F., Rolland, B., Zhang, X., Inoue, M., Matsuo, K., et al., 2011. Association between body-mass index and risk of death in more than 1 million Asians. *N Engl J Med* 364:719–729.
- Ziegler, J., 2007. Philosophers and physicians on the scientific validity of Latin physiognomy, 1200-1500. *Early Sci Med* 12:285–312.