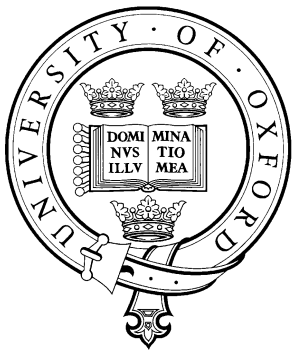
FETAL SKELETAL IMAGING USING 3D ULTRASOUND AND THE IMPACT OF MATERNAL VITAMIN D

Dr Christos Ioannou

Ptychion Iatrikes 1998, University of Athens, Greece



Merton College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

to the Faculty of Medical Sciences, University of Oxford.

Hilary Term 2012

FETAL SKELETAL IMAGING USING 3D ULTRASOUND AND THE IMPACT OF MATERNAL VITAMIN D

Dr Christos Ioannou
Merton College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy,
to the Faculty of Medical Sciences, University of Oxford.

Hilary Term 2012

Abstract

BACKGROUND Previous research suggests that vitamin D deficiency during pregnancy may be associated with suboptimal fetal growth, but direct evidence is lacking. Our objectives were 1) to develop a method for measurement of the fetal sphenoidal fontanelle area (FA) and femur volume (FV) using 3D ultrasound; 2) to create normal charts for FA and FV; and 3) to correlate FA and FV with maternal vitamin D concentration.

METHODS FA measurement in 3D was evaluated in vitro and in vivo. Different segmentation methods for FV measurement were explored. A novel FV method was described which consists of three linear measurements and a volume equation; this was validated in vitro and also by comparing FV measured sonographically to the true volume assessed by computed tomography (CT), in 6 cases following pregnancy termination. A cohort of 868 uncomplicated pregnancies was selected on the basis of strict inclusion criteria; participants underwent serial ultrasound scans for FV and multilevel modeling was used for the creation of a “prescriptive” FV chart. Finally, a different cohort of 357 healthy pregnant women had serum vitamin D levels and FV ultrasound at 34 weeks gestation and dual emission x-ray absorptiometry (DEXA) of their neonates in order to investigate the prenatal determinants of fetal bone mass.

RESULTS FA measurement was accurate in vitro, but unreliable in vivo and was therefore abandoned. A novel FV method had excellent agreement with CT and superior repeatability compared with segmentation-based methods. A normal FV chart was created and the regression equations for the median and percentile values were presented. Vitamin D demonstrated a significant correlation with FV.

CONCLUSIONS FV is a reliable sonographic marker of skeletal growth. Maternal vitamin D deficiency is associated with reduced FV. This finding has public health implications as reduced bone mass may increase the lifetime risk of osteoporosis, through fetal programming.

Acknowledgements

I would like to thank my supervisors A Papageorghiou (Senior Fellow in Fetal Medicine) and K Javaid (Lecturer in Metabolic Bone Disease) for their guidance; without their support this dissertation would not be possible. This research fellowship was supported by Prof S Kennedy (Head of the Nuffield Department of Obstetrics & Gynaecology); Prof A Noble (Professor of Biomedical Engineering), Prof J Villar (Chief Investigator of the INTERGROWTH 21st study) and Prof C Cooper (Professor of Rheumatology and Director of the MRC Epidemiology Resource Centre).

I am grateful to Dr I Sarris (Specialist trainee in Obstetrics & Gynaecology) who assisted me in performing ultrasound examinations and measurements for reproducibility analyses; Dr M Yaqub (Postdoctoral Fellow in Biomedical Imaging) who conceived the femur fusion technique and provided invaluable technical support in the use of biomedical imaging software; Dr P Mahon (Sonographer) for giving me access to 3D ultrasound examinations she performed for the Southampton Women's Survey; and the research staff at INTERGROWTH 21st, including recruitment midwives F Roseman and A Laister; and E Ohuma (Statistician).

I am indebted to Mr L Impey, Dr R Black and Dr P Chamberlain (Consultants in Fetal Medicine) for teaching me advanced fetal ultrasound; and my fellow colleague doctors R Napolitano, C Knight, A Brooks and S Collins who shared with me tips, advice and support. The CT validation study was facilitated by J Scott (Research Manager at the Stillbirth and Neonatal Death Society); the midwives at Prenatal Diagnosis and Level 7 of the John Radcliffe Hospital; and also the staff at the Radiology Department, including H Nichol (Radiographer) and Dr K Platt (Consultant Radiologist).

Finally, I am grateful to my father I Ioannou for his help in mathematical aspects of volumetric modeling and my family for their patience and support throughout my 3-year research fellowship.

Funding

This research fellowship was supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the Department of Health NIHR Biomedical Research Centres funding scheme. Philips Healthcare provided the HD9 ultrasound machine and technical assistance.

The INTERGROWTH 21st project was supported by a grant from the Bill and Melinda Gates Foundation to the University of Oxford (Grant ID# 49038).

The Southampton Women's Survey was funded from the Medical Research Council.

Ethics

All studies and analyses included in this thesis were conducted in the UK, with the participation of NHS patients and were approved by Ethical Committees within the National Research Ethics Service (NRES).

Details of the Committees and their approvals were as follows:

1. Postmortem fetal imaging with three dimensional ultrasound and computed tomography; Oxfordshire Research Ethics Committee B - Ref 09/H0605/124
2. The International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH 21st); Oxfordshire Research Ethics Committee C - Ref 08/H0606/139
3. Southampton Women's Survey; Southampton and South West Hampshire Local Research Ethics Committee – Ref 329/00

Table of contents

ABSTRACT 2

ACKNOWLEDGEMENTS 4

FUNDING 5

ETHICS 6

TABLE OF CONTENTS 7

LIST OF FIGURES 13

LIST OF TABLES 18

LIST OF ABBREVIATIONS 21

1 BACKGROUND 23

1.1 Principles of 2D and 3D ultrasound in pregnancy 23

1.1.1 Introduction 23

1.1.2 Ultrasound safety 23

1.1.3 Principles of 3D ultrasonography 24

1.1.4 Offline 3D image analysis 26

1.1.5 Applications in obstetrics 27

1.1.6 Ultrasonographic assessment of the fetal skeleton 29

1.2 A review of fetal volumetry: the need for standardization and definitions in measurement methodology 31

1.2.1 Introduction 31

1.2.2 Methods 32

1.2.3 Results and discussion 33

1.2.4 Conclusions 40

1.3 Fetal bone development and the role of vitamin D 40

1.3.1 Endomembranous bone differentiation 41

1.3.2 Endochondral bone differentiation 42

1.3.3 The growth plate 44

1.3.4 The role of vitamin D 45

1.3.5 Subclinical vitamin D insufficiency: a modern epidemic 46

1.3.6 Vitamin D insufficiency during pregnancy 48

1.3.7 Fetal effects of vitamin D insufficiency 49

1.3.8 Vitamin D supplementation in pregnancy: fetal effects 51

1.4 Background summary, hypotheses and aims 53

1.4.1 Hypotheses 54

1.4.2 Aims of the thesis 54

1.5 Publications and contributions 54

2 SPHENOIDAL FONTANELLE SURFACE AREA 55

2.1 In vitro validation of surface area measurement using a phantom 55

2.1.1 Introduction 55

2.1.2 Methods 56

2.1.3 Results 59

2.1.4 Discussion 62

2.1.5 Conclusions 64

2.2 Sphenoidal fontanelle measurement is not reliable in vivo 64

2.2.1 Introduction 64

2.2.2 Methods 65

2.2.3 Results 67

2.2.4 Discussion 69

2.2.5 Conclusions 71

2.3 Publications and contributions 71

3 FEMUR VOLUME MEASUREMENT USING MULTIPLANAR SEGMENTATION 73

3.1 Longitudinal segmentation does not improve measurement agreement 73

3.1.1 Introduction 73

3.1.2 Methods 74

3.1.3 Results 75

3.1.4 Discussion 77

3.1.5 Conclusions 78

3.2 Fusion of multiple volumes: a novel technique which can improve boundary definition 78

3.2.1 Introduction 78

3.2.2 Methods 79

3.2.3 Results 80

3.2.4 Discussion 83

3.2.5 Conclusions 84

3.3 Publications and contributions 85

4 A NEW SIMPLE METHOD FOR FEMUR VOLUME MEASUREMENT 86

4.1 In vitro validation of linear measurements on reconstructed planes 86

4.1.1 Introduction 86

4.1.2 Methods 87

4.1.3 Results 89

4.1.4 Discussion 91

4.1.5 Conclusions 92

4.2 A new method for femur volume measurement 92

4.2.1 Introduction 92

4.2.2 Methods 93

4.2.3 Results 97

4.2.4 Discussion 104

4.2.5 Conclusions 106

4.3 Publications and contributions 106

5 A LONGITUDINAL STUDY OF NORMAL FEMUR VOLUME 108

5.1 Introduction 108

5.2 Methods 108

5.2.1 Study design 108

5.2.2 Measurement technique 110

5.2.3 Statistical analysis 110

5.3 Results 112

5.4 Discussion 117

5.5 Conclusions 119

5.6 Publications and contributions 119

6 THE EFFECT OF MATERNAL VITAMIN D CONCENTRATION ON FETAL BONE 120

6.1 Introduction 120

6.2 Methods 121

6.3 Results 122

6.4 Discussion 128

6.5 Conclusions 131

6.6 Publications and contributions 131

7 DISCUSSION 133

7.1 Summary of main findings and discussion 133

7.2 Strengths 136

7.3 Limitations 137

7.4 Conclusions and implications for future research 139

8 APPENDIX 141

8.1 Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size 141

8.1.1 Introduction 141

8.1.2 Methods 142

8.1.3 Results 147

8.1.4 Discussion 157

8.1.5 Conclusion 163

8.1.6 Contributions 163

8.2 Publications arising from this thesis 164

8.2.1 Publications in peer-reviewed journals 164

8.2.2 Published abstracts from scientific conferences 164

9 REFERENCES 166

List of figures

Figure 1 Multiplanar imaging of the fetal head at the level of the biparietal diameter; transverse (A), coronal (B) and sagittal (C) planes 26

Figure 2 Fetal head visualized with surface and maximum rendering 27

Figure 3 Cross-sectional view of the femur (a) and its acoustic shadow (arrow). Speculative tracing of the inferior border (b) may reflect the true situation more closely, whereas tracing of the visible echo only (c) will underestimate the total volume 37

Figure 4 Endochondral ossification and the growth plate 43

Figure 5 Meta analysis of trials of Vitamin D supplementation in pregnancy, using birthweight as the outcome measure: all studies included 52

Figure 6 Meta analysis of trials of Vitamin D supplementation in pregnancy, using birthweight as the outcome measure: only studies at low risk of bias included 53

Figure 7 A flat vinyl tile with four holes of perfect geometrical shapes was scanned in a test tank by two sonographers, in four different depths and using two sweep directions: along the x axis and then along the y axis 57

Figure 8 Appearance and measurement of the phantom on rendered 3D ultrasound 58

Figure 9 Bland-Altman plots of interobserver percentage differences when tracing the same scan image (A) and on a newly acquired image (B); the solid line represents the mean; dotted lines show ± 2 standard deviations 60

Figure 10 Validity expressed as the percentage difference between the measured and true surface area versus shape (A) and depth (B); box plots demonstrate median, interquartile range, outliers (o) and extreme values (∗); p value calculated using the Kruskal Wallis test 61

Figure 11 Validity expressed as the percentage difference between the measured and true surface area versus sonographer (A) and direction of 3D sweep (B); box plots

demonstrate median, interquartile range, outliers (o) and extreme values (∗); p value calculated using the Mann Whitney U test 61

Figure 12 A transverse head volume acquired at the level of BPD is ideal for visualization of the sphenoidal fontanelle 65

Figure 13 Appearance and measurement of the sphenoidal fontanelle area on rendered 3D ultrasound 66

Figure 14 Bland-Altman plots of intraobserver percentage differences when measuring twice the same image (A) and when re-measuring an image from a newly acquired volume (B); the solid line represents the mean; dotted lines show ± 2 standard deviations 68

Figure 15 Technique for cross-sectional segmentation; the bone contour is traced on plane B, on several parallel slices; slices are then combined to calculate the volume 75

Figure 16 Technique for longitudinal segmentation; the bone contour is traced on plane A, on several parallel slices; slices are then combined to calculate the volume 75

Figure 17 Bland-Altman plots of intraobserver percentage differences for cross-sectional segmentation (A) and longitudinal segmentation (B); the solid line represents the mean; dotted lines show ± 2 standard deviations 76

Figure 18 Bland-Altman plots of interobserver percentage differences for cross-sectional segmentation (A) and longitudinal segmentation (B); the solid line represents the mean; dotted lines show ± 2 standard deviations 76

Figure 19 Six volumes of the entire femur were obtained from different angles and were fused using image registration and wavelet-based fusion 80

Figure 20 Images of the same femur before and after fusion; multiple ultrasound fields were combined into one fused image (arrow) 81

Figure 21 Bland-Altman plots of intraobserver percentage differences for segmentation on the single femoral view (A) and on the fused view (B); the solid line represents the mean; dotted lines show ± 2 standard deviations 82

Figure 22 Bland-Altman plots of interobserver percentage differences for segmentation on the single femoral view (A) and on the fused view (B); the solid line represents the mean; dotted lines show ± 2 standard deviations 82
Figure 23 Wire phantom and visualization of targets on the reference plane 88
Figure 24 Measurement protocol for linear measurements on the reference image..... 88
Figure 25 Wire phantom and visualization of targets on the reconstructed plane 89
Figure 26 Measurement protocol for linear measurements on the reconstructed image .. 89
Figure 27 Reconstructed image demonstrates misalignment (arrow) of the wire targets within a short distance (3 cm) from the ultrasound transducer 90
Figure 28 Bland-Altman plots of intraobserver measurement differences on the reference (A) and the reconstructed image (B); the solid line represents the mean; dotted lines show ± 2 standard deviations..... 90
Figure 29 The femur in the three constituent imaging planes: prior to measurement the orthogonal orientation axes (shown in green, red and blue) are aligned with the longitudinal axis of the femur in planes A and C and their intersection positioned at the mid-shaft (left A, B, C). The coronal A plane is then used for FL measurement (right A) and the sagittal C plane is used for PMD and MSD measurement (right C) 94
Figure 30 Schematic representation of the femur as a hyperboloid; MSD = mid-shaft diameter, PMD = proximal metaphyseal diameter and FL = femur length 95
Figure 31 Phantom fetus and the femur in 3D ultrasound..... 97
Figure 32 Bland-Altman plots of intraobserver percentage differences using the mathematical femur model (A) and the multiplanar segmentation technique (B); the solid line represents the mean; dotted lines show ± 2 standard deviations..... 99
Figure 33 Bland-Altman plots of interobserver percentage differences using the mathematical femur model (A) and the multiplanar segmentation technique (B); the solid line represents the mean; dotted lines show ± 2 standard deviations..... 99
Figure 34 Two representative cases: rendered CT images at 28 and 24 weeks respectively following pregnancy termination 101

Figure 35 CT images of the fetal femur in six cases following pregnancy termination; the cross-sectional outline of the proximal metaphysis (first column) and of the mid-shaft (second column) are near circular; the outline of the distal metaphysis (third column) is flattened and ovoid 102

Figure 36 Longitudinal femoral curvature in four representative cases: the curvature is more prominent at the medial femoral border..... 103

Figure 37 Distal metaphysis is rather flattened and ovoid: four representative cases 103

Figure 38 No movement artifact and fine artifact causing “fuzzy edges” 105

Figure 39 Scatter plot of the square root transformed data with fitted 5th, 50th and 95th percentile curves following multilevel modeling (A) and plot of SDS scores (B) demonstrating goodness of fit 113

Figure 40 Fetal femur volume charts: scatter plot with superimposed 5th, 50th and 95th percentiles (A) and chart demonstrating 3rd, 5th, 10th, 50th, 90th, 95th and 97th percentile curves (B) 115

Figure 41 Fetal femur volume charts showing the 10th, 50th and 90th centiles from the present study (solid lines) compared to those from the study by Chang et al (dotted lines) 118

Figure 42 The effect of maternal vitamin D concentration on femur length (A), proximal metaphyseal diameter (B), mid shaft diameter (C) and femur volume (D); all ultrasound measurements are adjusted for gestational age and FV axis is shown in logarithmic scale; box plots demonstrate mean values; error bars show the 95% confidence intervals; p = statistical significance from univariate linear regression analysis; r = Pearson correlation coefficient 125

Figure 43 Linear regression scatterplots showing femur length (A), proximal metaphyseal diameter (B), mid shaft diameter (C) and femur volume (D) vs. maternal vitamin D concentration; all ultrasound measurements are adjusted for gestational age and FV axis is shown in logarithmic scale; p = statistical significance from univariate linear regression analysis; r = Pearson correlation coefficient..... 126

Figure 44 Consort flow diagram of literature assessment 143

Figure 45 Methodological quality of included studies: Study design 153

Figure 46 Methodological quality of included studies: Statistical methods 154

Figure 47 Methodological quality of included studies: Reporting methods 154

Figure 48 Exclusion criteria used in included studies..... 155

Figure 49 Ultrasound quality assurance measures in included studies..... 156

Figure 50 Use of presentation methods in included studies..... 157

List of tables

Table 1 Volumetry studies of fetal organs with normal values at representative and comparable gestational ages; all volumes quoted in ml, n = number of fetuses, GA = gestational age in weeks; for each organ one study was used as reference (author underlined) 34

Table 2 Reasons for poor inter-researcher agreement in volumetry 36

Table 3 Shapes and their true surface areas calculated using standard mathematical equations; S = surface area, t = side length, h = height, b = base 58

Table 4 Repeatability of surface area measurement in vitro; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval 60

Table 5 Feasibility of sphenoidal fontanelle area measurement at different gestational ages; GA = gestational age, n= number of cases 67

Table 6 Intraobserver repeatability of sphenoidal fontanelle area measurement in vivo; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval 68

Table 7 Repeatability of cross-sectional vs. longitudinal segmentation; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval 77

Table 8 Repeatability of femur volume using cross-sectional segmentation on single vs. fused datasets; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval 83

Table 9 Repeatability of linear measurement on the reference and the reconstructed image; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval 91

Table 10 Volume measurements of a phantom femur using the mathematical model and segmentation technique 98

Table 11 Repeatability of femur volume measurement using the mathematical model and multiplanar segmentation; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval 100

Table 12 Characteristics of cases included in the CT validation study 104

Table 13 Demographic characteristics of study participants; data are presented as mean (SD) or n (%) 112

Table 14 Equations of femur volume (FV = ml) as a function of gestational age (w = weeks) 114

Table 15 Fitted median and percentile FV values (in ml) following multilevel modeling; GA = gestational age in weeks +0 days 116

Table 16 Maternal and fetal characteristics; height, weight and BMI measured prior to conception; smoking and alcohol consumption documented in early pregnancy; exercise, triceps skinfold thickness, and vitamin D levels assessed in late gestation; FL, PMD, MSD, FV measured by 3D ultrasound in late gestation; BMC, BMD assessed by dual emission x-ray absorptiometry within two weeks following birth .. 123

Table 17 Univariate effect of maternal characteristics on fetal ultrasound measurements; p value (statistical significance) and r coefficient (Pearson correlation coefficient) derived from linear regression analysis; for categorical variables (*) the p value is derived from the independent samples t-test; FL, PMD, MSD and logFV are adjusted for gestational age at ultrasound 124

Table 18 Univariate effect of maternal characteristics on neonatal measurements; p value (statistical significance) and r coefficient (Pearson correlation coefficient) derived from linear regression analysis; for categorical variables (*) the p value is derived from the independent samples t-test; BW is adjusted for gestational age at delivery; BMC, BMD are adjusted for gestational age at delivery and postnatal age at DEXA 127

Table 19 Search strategy 144

Table 20 List of methodological quality criteria; GA = gestational age, SD = standard deviation, LMP = last menstrual period, CRL = crown rump length; for “LMS method”

see Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr 1990; 44: 45-60 145

Table 21 Characteristics of included studies (CM: cross-sectional data, several scans per fetus, C1: cross-sectional data, one scan per fetus, CS: cross-sectional data, number of scans per fetus not sated, LC: longitudinal data with cross-sectional analysis, LL: longitudinal data with longitudinal analysis) 148

Table 22 Dating methods (LMP: last menstrual period; US: ultrasound; CRL: crown rump length) 155

Table 23 Comparison of measurement values amongst highest scoring studies 162

List of abbreviations

2D	Two dimensional
3D	Three dimensional
AC	Abdominal circumference
AIUM	American Institute of Ultrasound in Medicine
ALARA	As Low As Reasonably Achievable
BMC	Bone mineral content
BMD	Bone mineral density
BMI	Body mass index
BMUS	British Medical Ultrasound Society
BPD	Biparietal diameter
C1	Cross-sectional data, one scan per fetus
CI	Confidence interval
CM	Cross-sectional data, several scans per fetus
CRL	Crown-rump length
CS	Cross-sectional data, number of scans per fetus not stated
CT	Computed tomography
CV	Within-subject coefficient of variation
DBP	Vitamin D binding protein
DEXA	Dual emission x-ray absorptiometry
FA	Sphenoidal fontanelle area
FGFR	Fibroblast growth factor receptor
FL	Femur length
FV	Femur volume
GA	Gestational age
HC	Head circumference

HOX	Homeobox containing gene
ICC	Intraclass correlation coefficient
IGF	Insulin-like growth factor
IHH	Indian hedgehog homolog gene
ISUOG	International Society of Ultrasound in Obstetrics and Gynaecology
LC	Longitudinal data with cross-sectional analysis
LL	Longitudinal data with longitudinal analysis
LMP	Last menstrual period
MI	Mechanical index
MRI	Magnetic resonance imaging
MSD	Mid-shaft diameter
MSX	Muscle-segment homeobox related gene
PC	Personal computer
PMD	Proximal metaphyseal diameter
PTH	Parathyroid hormone
PTHrP	Parathyroid hormone-related peptide
QUS	Quantitative ultrasound
RCT	Randomized controlled trial
SD	Standard deviation
SDS	Standard deviation score
SHH	Sonic hedgehog homolog gene
TI	Thermal index
TSF	Triceps skinfold thickness
UVB	Ultraviolet radiation B
VDR	Vitamin D receptor
VOCAL	Virtual organ computer-aided analysis
WHO	World Health Organization

1 Background

Assessment of fetal skeletal growth during pregnancy is important since intrauterine bone growth is associated with postnatal and adult skeletal health, through fetal programming. The following chapter has three objectives: 1) to provide a summary of the available evidence regarding the use of obstetric ultrasound in assessing the fetal skeleton; 2) to describe known associations between fetal bone growth and maternal vitamin D metabolism; and 3) to summarize the hypotheses and objectives of this thesis.

1.1 Principles of 2D and 3D ultrasound in pregnancy

1.1.1 Introduction

Two dimensional (2D) ultrasound been used extensively in the assessment of fetal wellbeing, since it was introduced in the 1960s. It is overall a safe, inexpensive and accurate imaging modality[1]; it can provide real-time, bedside diagnosis; it generates valuable information on the size, anatomical integrity and cardiovascular physiology of the fetus; it aids bonding between the mother and her expected baby; and by establishing the diagnosis of fetal conditions, it allows medical intervention: this may include intrauterine medical or surgical treatment, planned delivery for timed postnatal intervention or termination of pregnancy.

1.1.2 Ultrasound safety

The interaction between ultrasound and human tissue has the potential for thermal and non-thermal effects. Thermal effect refers to tissue temperature elevation as a result of exposure to ultrasound; mechanical effect or cavitation describes the movement, expansion or collapse of gas bubbles within tissues, which has the potential for cellular damage[2]. The magnitude of those effects depends on the ultrasound emission settings,

type of ultrasound (B-mode or Doppler) and the duration of continuous exposure. During an examination, ultrasound scanners should display on-screen the mechanical (MI) and thermal index (TI): these indices provide a safeguard against potential biologic effects from the given machine settings. A safety statement has been issued by the International Society of Ultrasound in Obstetrics and Gynaecology (ISUOG)[3]; safety guidelines have also been published by the American Institute of Ultrasound in Medicine (AIUM)[4] and the British Medical Ultrasound Society (BMUS)[5] ; these recommend that exposure time and acoustic output settings should be kept to the lowest levels consistent with obtaining diagnostic information - the ALARA principle ("As Low As Reasonably Achievable").

Despite those potential effects at tissue level, it is important to emphasize that ultrasound examination is safe for the mother and fetus and that it has never been associated with adverse clinical outcomes. A recent systematic review of 61 observational and interventional studies on behalf of ISUOG and the World Health Organization (ISUOG – WHO) Fetal Growth Study Group concluded that prenatal ultrasound is not linked with any maternal or perinatal complications, impaired physical or neurological development, increased risk for malignancy in childhood, subnormal intellectual performance or mental disease; there was only a weak association between exposure to ultrasound and non-right handedness in boys[6].

1.1.3 Principles of 3D ultrasonography

The emergence of 3D ultrasound systems was facilitated by the advent of computational and storage capacity of modern computers. The principle of ultrasonic image acquisition is in fact identical to 2D imaging; it is rather the ability to process a vast amount of digital imaging information which has made this technology available to every day clinical practice. A 2D image only contains information on the x and y coordinates. A 3D ultrasound tomogram is accomplished by the acquisition of several consecutive and near parallel 2D images, while the ultrasound transducer moves along the z coordinate. The imaging information is thus contained in a block, rather than a single plane; each point in

this block is called a voxel, rather than a pixel. The entire block of imaging data is saved in a memory medium, such as the ultrasound machine hard drive or transportable memory media.

There are three different types of 3D ultrasound systems: freehand systems, automated mechanical 3D transducers and matrix array transducers. A *free-hand 3D system* consists of a standard 2D ultrasound scanner with a motion sensor attached onto the transducer. Several planes of the region of interest are acquired while the operator is sliding the transducer manually along a certain direction. The motion sensor tracks the movement of the transducer and aids the correct spatial integration of the acquired 2D planes[7, 8].

The need for an external tracking sensor is obviated with the use of an *automated mechanical 3D probe*. This is a larger and heavier enclosure which contains a 2D transducer fitted with a rotating mechanism. A 3D image dataset is generated at the touch of a button: while the operator holds the probe still, the transducer rotates by a specified angle and collects several 2D planes. One such 3D sweep lasts 2-9 seconds, depending on the angle and speed of the sweep, both of which are adjustable. A larger angle or a slower sweep speed result in slower acquisition and a larger image data file. However if the region of interest is moving, a slow sweep speed is detrimental because of movement artifact; for instance when the fetus is very active, a faster 3D sweep may be preferable in order to minimize artifact.

Finally, the newest generation of 3D probes is the *matrix array transducer*. Where standard 2D probes are made of a linear array of piezoelectric crystals capable of scanning in one plane only, matrix array probes have several parallel lines of crystals which can all “fire” ultrasound simultaneously; this generates a real-time 3D scan as it collects imaging information from all three coordinates simultaneously. Much more elaborate and expensive to manufacture, such transducers are rarely used in obstetrics. They are however valuable when studying fast moving structures, like in 3D echocardiography[9] where they can minimize imaging artifact.

1.1.4 Offline 3D image analysis

After acquisition, the 3D volume is displayed on-screen using one of two different modalities: multiplanar navigation or volume rendering. With *multiplanar navigation* any of the three perpendicular planes which are intersecting over a particular point of the image dataset can be displayed. For instance, if the region of interest is the fetal head and the volume was originally obtained from a transverse plane at the level of the thalami, then the three orthogonal planes will be the transverse, coronal and the sagittal (Figure 1). By convention, these planes are called plane A - the original plane at the time of acquisition - and planes B and C which are the reconstructed planes, at 90^{o} angle to the original. Any plane can be selected and different slices can be viewed, by scrolling along the third coordinate; for instance the transverse view of the head can be scrolled from the top of the skull down to its base. Alternatively, it is possible to reconstruct oblique planes at a varying angle from the A plane. Any reconstructed plane is displayed as a grey-scale 2D image; this is indistinguishable from any other 2D scan image and allows conventional distance and ellipse measurements.

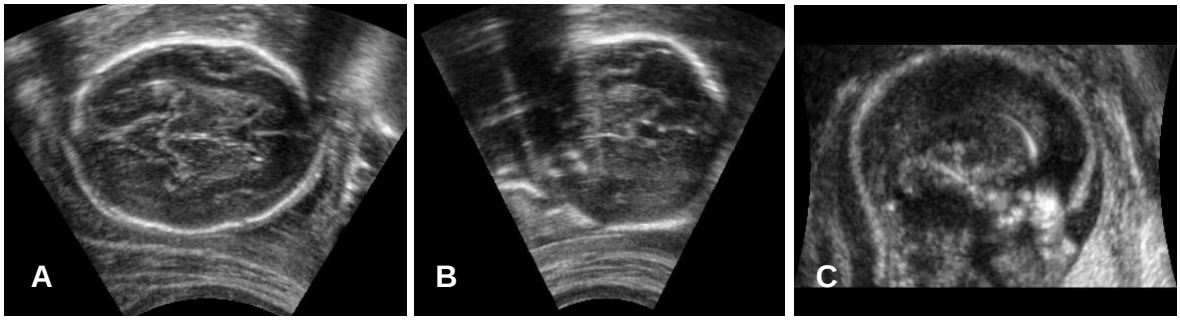


Figure 1 Multiplanar imaging of the fetal head at the level of the biparietal diameter; transverse (A), coronal (B) and sagittal (C) planes

When using *volume rendering,* the entire image dataset is displayed as a single image which has photographic effects like lighting and depth. This is an artificial picture, where points from different planes are projected onto a single visualisation plane. It is not routinely used for measurement purposes, but it has life-like appearance and it is more

easily recognizable by patients. There are two main types of rendering: surface and maximum rendering. In *surface* rendering the voxels which are reflected first are displayed in preference; this technique is therefore excellent for displaying skin surfaces and is usually applied in order to obtain images of the fetal face. In *maximum* rendering, the points of maximum intensity are only displayed and therefore this modality is used to demonstrate skeletal structures (Figure 2).

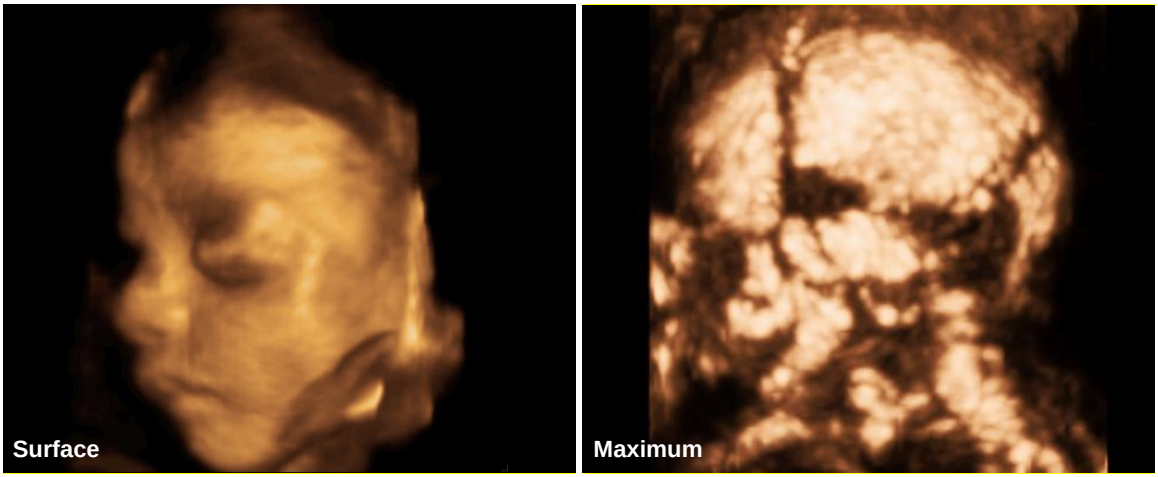


Figure 2 Fetal head visualized with surface and maximum rendering

1.1.5 Applications in obstetrics

Although there have been numerous studies of 3D ultrasound in obstetrics since the mid-1990s, real-time 2D ultrasound remains the standard of care in pregnancy worldwide and one of the most commonly practiced medical investigations. 3D is widely available in commercial, non-clinical settings for the generation of rendered images and movies of the developing fetus. Rendered 3D images of a fetus are more easily recognizable by mothers[10]. However their emotional impact is disputed, with some studies claiming improved parental bonding[11] and others demonstrating no effect whatsoever[10, 12].

In clinical practice, 3D can be used as an adjunct to 2D scanning in order to confirm or better characterize fetal anomalies[13]. The entire fetus can be accommodated in a single volume sweep only during the first trimester of pregnancy[14]. Beyond 14 weeks the fetus is

too large for a single sweep and it is common practice to acquire separate anatomical compartments i.e. head, chest, abdomen, limbs etc. There are several case reports and small series describing fetuses with abnormalities which were demonstrated on 3D examination, after the anomalies were diagnosed and described in 2D[15, 16]. It is not surprising that 3D can demonstrate virtually any fetal abnormality which is detectable on 2D ultrasound[17], because the technology of ultrasound emission and acquisition is identical. It has also been claimed that a full anatomic survey can be performed with offline manipulation of saved 3D volumes and that this can take less time than a real time 2D anomaly scan[18]. However, adequate views for a complete "offline" 20 week scan can only be obtained in 66% of cases[18]; or as little as 33% of cases for a basic anatomic survey at 13 weeks gestation[19]. Furthermore, it has not been assessed whether an offline 3D survey is as reliable in identifying fetal anomalies as the routine 20 week scan. In other words, while 3D may be a useful adjunct, it cannot at this stage substitute 2D scanning in anomaly screening.

Saved 3D volumes can also be used in order to perform offline fetal biometry in a fashion similar to 2D biometry. Using offline multiplanar navigation, standard planes for measurement of the head, abdomen and femur can be generated and conventional measurements, such as distance or circumference, can be performed. It has been demonstrated that head circumference, biparietal diameter, abdominal circumference and femur length measured offline on a 3D volume are almost identical to the real-time 2D measurements[20]. Intraobserver and interobserver repeatability of measurements after volume manipulation are comparable[21] or better[20] than the repeatability of real-time 2D biometry by an experienced sonographer. The underlying assumption however is that a fully trained ultrasonographer has acquired the original volume scans; and that the orientation of fetal parts during scan acquisition is close to the standard plane orientation for 2D biometry (i.e. horizontal femur for a femur length, or horizontal falx cerebri for head circumference). It should not be inferred that offline biometry is feasible or reliable when the 3D volumes are performed in a random fashion by a non-trained operator.

Novel offline measurements can be performed on reconstructed imaging planes other than the standard planes for fetal biometry. For instance, measurement of the fronto-maxillary facial angle was described on a reconstructed sagittal view of the fetal face at 12 weeks[22]; also the fetal iliac angle from a transverse pelvic plane at 21 weeks gestation[23]. Although such novel planes could be obtained with real-time 2D scanning, the use of saved 3D volumes obviates the need for women to be subjected to prolonged ultrasound examinations and it therefore facilitates research.

1.1.6 Ultrasonographic assessment of the fetal skeleton

The principle of ultrasonic image generation is that when high frequency sound waves traverse the human body, they are reflected back from any point where they meet tissues of different density. The transducer emits several ultrasound waves per second, then collects the reflected waves and transforms them into electrical signal. The ultrasound scanner uses information from the intensity of the reflected ultrasound and the depth of reflexion in order to recreate a 2D image of those structures that crossed the path of the beam.

The distinct ultrasonic properties of fetal bone pose unique challenges when scanning the fetus. Due to its high reflectivity, bone is easy to identify as it generates a bright, strong echo. At the same time the ultrasound beam becomes much attenuated, as a considerable fraction of its energy gets deflected. This becomes evident on screen as an area of shadow, immediately underneath any skeletal structure, known as acoustic shadow.

Femur length measurement in 2D is not affected by acoustic shadowing as the diaphyseal ends are displayed clearly enough for calliper application. As a result, intraobserver and interobserver repeatability of femur length is good[24]. However, acoustic shadowing may pose a considerable problem when trying to delineate the femur shape in all three dimensions; for instance when trying to calculate the femur volume in 3D. A methodology

for femur volume measurement has been described[25] and will be considered in greater length in section 1.2.

Another area of interest in 3D is the fetal skull. Multiplanar navigation of a head volume can be used qualitatively in order to ascertain presence or absence of intracranial structures[26, 27]. It can also aid the quantitative assessment of the facial profile[22, 28], skull base[29] and mandible[30] by performing simple distance and angle measurements. More complex geo-morphometric models have been applied in order to describe the curvature of the cranial vault in normal and abnormal fetuses[31]. A number of publications have investigated the use of rendered imaging in order to enhance the recognition of facial clefts, especially those of the soft and hard palate[32, 33]. The use of maximum rendering can enhance the visualization of sutures and fontanelles as early as 16 weeks[34] and throughout the second half of gestation[35]. The development of the metopic suture has been studied from 9 to 34 weeks both in normal fetuses[36] and fetuses with select craniofacial anomalies[37]. In addition, it has been argued that the ultrasonographic visualization of fontanelles may aid the diagnosis of craniosynostosis syndromes[38, 39].

For successful demonstration of a suture, the 3D transducer should be positioned immediately over the suture of interest. A feasibility study[35] concluded that the visualization rate for lateral sutures (coronal suture, sphenoidal fontanelle) is in excess of 92%; whereas for midline sutures (sagittal, posterior fontanelle) it ranges from 42 – 82%. This lower visualization rate for midline sutures is attributed to the fact that fetuses in late gestation are often in a low cephalic presentation and this makes the top of the skull less accessible.

Rendered 3D ultrasound images have been used in order to perform surface area measurement of the fetal anterior fontanelle: reference charts for anterior fontanelle size have been published[40] and an association between Downs syndrome and larger anterior fontanelles has been reported[41]. However, measurement on rendered 3D images should be interpreted with caution as this technique has not been validated; while volume calculation is very reliable in vitro[42], it cannot be assumed that surface area measurement

on rendered 3D has the same reliability. Rendered images are artificial projections of ultrasound echoes from different depths and this has the potential for significant error. Such error could also result from caliper application if the demarcation between bone and suture is not clear; from distortion during reconstruction of the rendered image; or true size variation due to molding of skull bones or external abdominal pressure.

1.2 A review of fetal volumetry: the need for standardization and definitions in measurement methodology

1.2.1 Introduction

The ability to perform volume calculations using 3D ultrasound is an added, valuable tool over the standard fetal biometry. With 2D ultrasound, volume estimations of symmetric and near spherical structures can be achieved using three orthogonal diameters and then applying the volume formula of an ellipsoid shape. Although this gives a reasonable approximation for symmetrical structures such as spherical cysts, it is less useful for complex structures and shapes. The emergence of automated 3D ultrasound transducers has allowed for accurate measurement of volumes, even in organs that are of irregular or asymmetrical shape[43, 44].

Such volume calculation has now been described for a variety of fetal organs and structures. In the majority of cases this is accomplished by manual organ segmentation. This calculation is achieved with manual tracing of the organ's outline on several consecutive imaging planes, using one of two methods: the *multiple parallel plane method (multiplanar),* where the organ outline is traced manually on multiple parallel slices while scrolling from the one end of the organ to the other; and the *rotational Virtual Organ Computer-aided AnaLysis* (VOCAL, GE Healthcare, Zipf, Austria) where organ segmentation is performed on rotating slices which are intersecting along a chosen central axis. With VOCAL, each slice is generated following rotation by a fixed angle; thus, for a

rotation step of 30^{o}, only 6 slices need to be segmented; whereas with an angle of 9^{o}, a total of 20 planes will be generated.

Newer volumetric tools include the semi-automated, virtual reality system I-Space VR (Barco, Belgium)[45] and the automated segmentation tool for fluid filled spaces Sono-AVC (GE Medical Systems, Zipf, Austria)[46]. For certain organs, the volume of interest can either be traced directly, or it can be calculated by subtraction from a larger anatomical compartment: for instance lung volume can be calculated either by tracing both lung fields directly[47] or by subtracting the heart from the chest volume[48]; similarly first trimester embryo volumes can be measured directly[49] or by subtracting the amniotic fluid and yolk sac volumes from the gestational cavity[46].

Using these techniques, volume measurement has been described for the fetal brain[50, 51], cerebellum[52], lungs[53, 54], liver[55], kidneys[56], upper arm[57], thigh[58], humerus[59], femur[25], spinal canal[60]; also first trimester volumes of the entire fetus[14, 61] and placental volumes in the first[62] or second trimester[63]. It has been proposed that the entire thigh volume[64] or fractional thigh volume[65] can be incorporated into fetal weight estimation models, which predict birth weight more accurately than existing 2D formulas. Lung volume calculation may be a useful prenatal predictor of lung hypoplasia in fetuses at risk[66, 67]. It has also been proposed that placental volume estimation may help discriminate healthy pregnancies from those at risk of pre-eclampsia[68, 69], intrauterine growth restriction[69]; however it is not a useful predictor of fetuses with a-thalassemia[70] or chromosomal abnormalities[71].

The objective of this review was to compare reported normal volumes from different studies for the same fetal organ in order to highlight potential discrepancies and their magnitude; and to identify the reasons for such discrepancies.

1.2.2 Methods

We conducted a literature review with the aim of identifying all studies reporting fetal organ volumetry up to December 2010. We only included fetal organs for which more than

one study existed in the literature. In order to compare differences in measured volumes between different studies, a specific gestational age was chosen per organ. For each organ, the largest cross-sectional study was used as the “reference study”. For every other study, the difference between the reported median volume and the reference median was assessed. This difference was expressed both as a percentage and as a z-score; z-scores were calculated using the standard deviations (SD) from the reference study. The use of z scores is a practical tool for comparing a new biometric reference against an existing one, since it translates measurement differences into percentile discrepancies[72]. For example, if the difference between two studies gives a z-score in excess of 1.65, it means that the median volume reported by one study is above the 95th centile according to the reference study. Similarly a z-score of less than -1.65 suggests that the median volume reported by one study is below the 5th centile according to another.

1.2.3 Results and discussion

Considerable differences in reported “normal” values were found for brain, liver, lungs and femur volume, whereas smaller discrepancies and better agreement were noted for cerebellum and embryo volume (Table 1). It could be argued that some of the observed variations may be due to differences in population characteristics, risk profile, ethnic variations or changes over time. However, when differences are large, such demographic characteristics are unlikely to explain all the variation: for example, in the case of right kidney volume, there are two studies from the same country and the same year[73, 74], where the normal value differs by 1.74 SD at 34 weeks gestation; this means that a renal volume of 7.5 ml is the 50th percentile according the one study or the 4th percentile according to the other. Similarly, it is illogical to think that a difference as large as -2.94 SD for lung volume or 2.22 SD for brain volume could reflect true biologic variation.

Table 1 Volumetry studies of fetal organs with normal values at representative and comparable gestational ages; all volumes quoted in ml, n = number of fetuses, GA = gestational age in weeks; for each organ one study was used as reference (author underlined)

Author	Year	Country	n	Design	Method	GA	Volume	Difference	z -score
Brain									
Roelfsema NM[51]	2004	Netherlands	68	Longitudinal	VOCAL 30	28	177	-9%	-1.18
Chang CH[50]	2003	Taiwan	203	Cross-sectional	Multiplanar	28	194.59	0%	0.00
Endres LK[75]	2001	US	85	Cross-sectional	Multiplanar	28	227.79	17%	2.22
Cerebellum									
Hata T[76]	2007	Japan	13	Longitudinal	Multiplanar	32	8.2	-11%	-0.45
Chang CH[77]	2000	Taiwan	231	Cross-sectional	Multiplanar	32	9.25	0%	0.00
Araujo Junior E[78]	2007	Brazil	52	Longitudinal	VOCAL 30	32	10.99	19%	0.75
Rutten MJ[52]	2009	Netherlands	27	Longitudinal	Multiplanar	32	11.35	23%	0.91
Embryo									
Falcon O[14]	2005	UK	417	Cross-sectional	VOCAL 30	10	no data at 10 weeks		
Blaas H-GK[79]	2006	Norway	44	Cross-sectional	Multiplanar	10	4.24	-27%	-0.75
Aviram R[49]	2004	Israel	72	Cross-sectional	VOCAL 30	10	4.62	-21%	-0.57
Gadelha PS[80]	2006	Brazil	25	Longitudinal	VOCAL 6	10	5.06	-13%	-0.37
Araujo Junior E[81]	2010	Brazil	64	Cross-sectional	XI VOCAL	10	5.12	-12%	-0.34
Rousian M[45]	2010	Netherlands	42	Longitudinal	I-space VR	10	5.35	-8%	-0.23
Rolo LC[82]	2009	Brazil	50	Cross-sectional	VOCAL 30	10	5.44	-7%	-0.19
Bagratee JS[83]	2009	UK	166	Cross-sectional	Multiplanar	10	5.84	0%	0.00
Martins WP[84]	2009	Brazil	20	Longitudinal	VOCAL 15	10	6.2	6%	0.17
Femur									
Mahon P[85]	2009	UK	483	Longitudinal	Multiplanar	34	2.8	-27%	-2.55
Chang CH[25]	2007	Taiwan	304	Cross-sectional	Multiplanar	34	3.82	0%	0.00
Liver									
Chang CH[50]	2003	Taiwan	226	Cross-sectional	Multiplanar	32	54.57	0%	0.00
Chang FM[86]	1997	Taiwan	55	Cross-sectional	Multiplanar	32	62.02	14%	0.94
Kuno A[87]	2002	Japan	14	Longitudinal	Multiplanar	32	63.3	16%	1.11
Boito SME[88]	2002	Netherlands	85	Cross-sectional	Multiplanar	32	66.42	22%	1.50
Laudy JAM[89]	1998	Netherlands	25	Cross-sectional	Multiplanar	32	72.02	32%	2.21
Dos Santos Rizzi MC[90]	2010	Brazil	53	Longitudinal	Multiplanar	32	74	36%	2.46
Lungs									
D'Arcy TJ[91]	1996	UK	20	Cross-sectional	Multiplanar	28	17.2	-52%	-2.94
Hata T[92]	2007	Japan	14	Longitudinal	Multiplanar	28	24.78	-31%	-1.76
Osada H[93]	2002	Japan	125	Cross-sectional	Multiplanar	28	24.99	-31%	-1.73
Bahmaie A[94]	2000	UK	54	Longitudinal	Multiplanar	28	25.26	-30%	-1.69
Moeglin D[95]	2005	France	622	Cross-sectional	Multiplanar	28	26.81	-26%	-1.45
Araujo Junior E[96]	2008	Brazil	51	Cross-sectional	Multiplanar	28	30.8	-15%	-0.83
Chang CH[97]	2003	Taiwan	152	Cross-sectional	Multiplanar	28	32.95	-9%	-0.49
Pohls UG[47]	1998	Germany	56	Cross-sectional	Multiplanar	28	33.39	-8%	-0.42
Britto ISW[98]	2009	Brazil	50	Cross-sectional	VOCAL 30	28	35.83	-1%	-0.05
Peralta CFA[99]	2006	UK	650	Cross-sectional	VOCAL 30	28	36.12	0%	0.00
Ruano R[100]	2004	France	109	Cross-sectional	VOCAL 30	28	39.8	10%	0.57
Gerards FA[54]	2006	Netherlands	78	Longitudinal	Multiplanar	28	40.67	13%	0.71
Lee A[48]	1996	Austria	78	Mixed	Multiplanar	28	41.96	16%	0.91
Laudy JAM[101]	1998	Netherlands	29	Cross-sectional	Multiplanar	28	42.29	17%	0.96
Sabogal JC[102]	2004	US	75	Longitudinal	Multiplanar	28	43.3	20%	1.11
Right kidney									
Hsieh YY[73]	2000	Taiwan	112	Cross-sectional	Multiplanar	34	7.5	-37%	-1.74
Tedesco GD[103]	2009	Brazil	57	Longitudinal	VOCAL 30	34	11.7	-1%	-0.06
Yu CH[74]	2000	Taiwan	134	Cross-sectional	Multiplanar	34	11.86	0%	0.00

The majority of these studies have been published in the last decade, therefore true biologic change over this time period is unlikely; similarly, inter-racial factors may contribute to some size variability, but these should be small differences. Rather, such large variations are much more likely to be due to methodological inconsistencies.

While trying to explain the reasons for such a wide variability between studies, it becomes easily apparent that fetal volumetry is fraught with considerable methodological heterogeneity in the way that data are collected, analyzed and presented (Table 2). Although 3D ultrasound has been in clinical practice for several years, a universal standard for digital image data storage and analysis has yet to be established for 3D ultrasound systems. Commercially available ultrasound machines operate proprietary file formats that preclude viewing and analysis of 3D image datasets outside the specific machine and its bundled software tools. It is reasonable to expect that a common file interface and universally available measurement tools would promote consistency across investigators and would facilitate research. 3D ultrasound users should plea for such an agreement among manufacturers. It is true that technological advances in biomedical imaging and novel measurement techniques will keep emerging. It is equally essential that each new method is subject to robust validation, followed by continuous quality assurance thereafter.

An essential aspect of any clinical examination or research study of fetal biometry is the explicit definition of ultrasonographic image landmarks and caliper placement. This is especially true for studies of 2D biometry where, for instance, great effort is taken in defining the correct imaging plane for the head and abdominal circumference or the femur length. Image scoring tools have been devised in order to audit the adequacy of planar image measurement[104] and have been shown to improve training and consistency[105, 106]. In volumetric scanning, not just one but several planes need to be defined consistently. However, remarkably little is specified about measurement landmarks and correct caliper application. In 3D, errors in caliper placement will be multiplied over the volume. It must also be remembered that the constituent planes of the entire 3D scan suffer from at least

the same limitations as B-mode scanning, such as image speckle and acoustic shadowing.

Table 2 Reasons for poor inter-researcher agreement in volumetry

Heterogeneity of 3D ultrasound system platforms and methods
– File formats incompatible between different manufacturers – Different proprietary measurement software – Alternative measurement methods: direct vs. subtracted vs. automated
Inconsistent interpretation of ultrasonic image
– Unclear measurement planes and anatomical landmarks – Lack of consensus on tracing – Inherent ultrasonic artifacts (signal attenuation, speckle, shadowing) – Fetal movement artifact
Analysis and reporting of measurement error
– Lack of inter-observer repeatability data – Repeated tracing of the same 3D volume – Inappropriate statistical methods to assess repeatability
Lack of method validation
– In vitro validation reflects the technical performance of the ultrasound machine only – In vivo validation is practically difficult

An example of the limitations introduced by lack of definitions is in the methodology proposed for femur volume calculation. Briefly this involves the acquisition of a volume of the fetal thigh, using the longitudinal view of the femur as the reference plane. The femur is then traced on several parallel slices in the cross-sectional plane, from one diaphyseal end to the other. Although seemingly a straightforward technique, it becomes quite problematic in practice. A typical cross-sectional view of the femur can be seen in Figure 3. The superior outline of the bone is clear and easy to delineate, whereas the inferior part fades and disappears into the acoustic shadow. Assuming that the femoral shaft is nearly cylindrical, its cross-section should have the shape of a ring as shown in Figure 3b. Yet, it is impossible to ascertain the inferior femoral border and tracing of the bone is thus completely speculative. Alternatively, the visible femoral echo only can be traced as in Figure 3c, but this would lead almost certainly to underestimation of the femur volume.

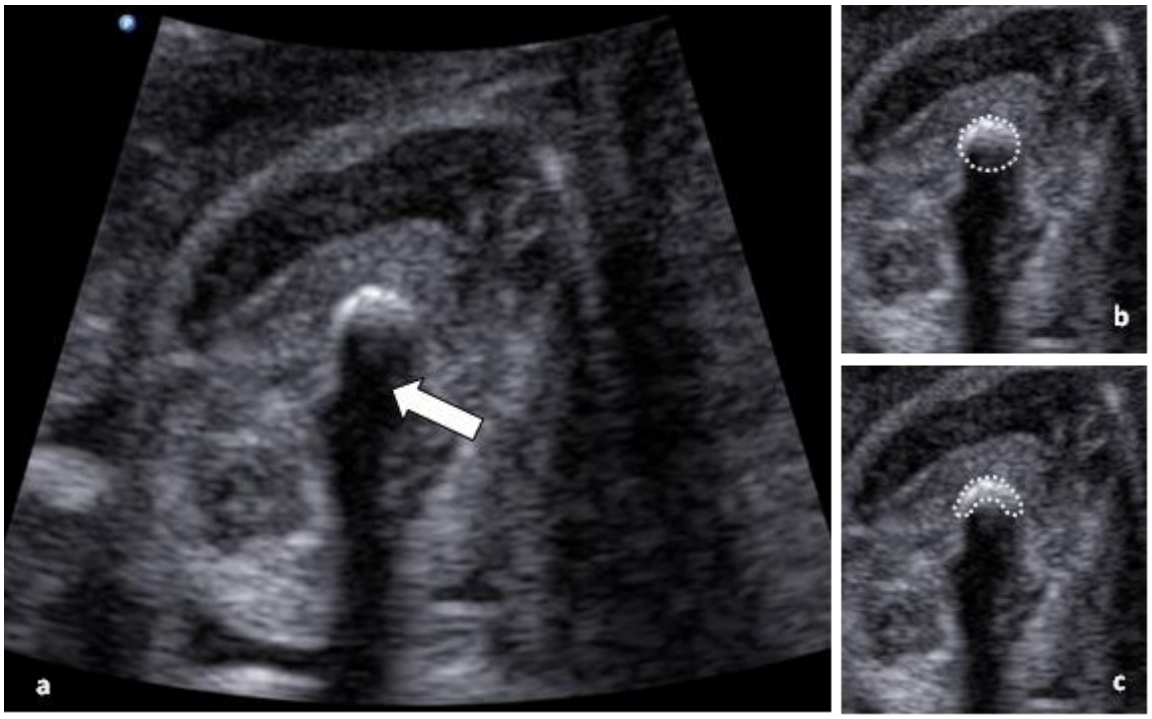


Figure 3 Cross-sectional view of the femur (a) and its acoustic shadow (arrow). Speculative tracing of the inferior border (b) may reflect the true situation more closely, whereas tracing of the visible echo only (c) will underestimate the total volume

There is no agreement about the correct way to segment the volume of the femur. Thus, femur volumetry using the multiplanar methodology has been reported, but in two recent studies[25, 85] there was a striking discrepancy in the median femur volumes: 3.82 ml vs. 2.80 ml at 34 weeks' gestation or a Z-score of −2.55 SD. This means that 2.8 ml is at the 50th centile according to one study[85] but below the 1st centile according to the other[25].

In order to assess measurement error, a minimum of two measurements should be obtained by the same observer (intraobserver error); also at least one additional measurement by a different observer (interobserver error). Studies carried out by a single operator have limited generalizability as they do not account for interobserver error. In spite of high intraobserver agreement, a measurement method may still be inconsistent when performed by different investigators. Repeatability reports can also be hampered by the choice of statistical methods used in the analysis of repeat measurement data. For

instance the use of paired t-test comparisons may highlight statistically significant measurement differences that are clinically insignificant. Conversely, a linear regression analysis may demonstrate excellent correlation between two observers, but correlation does not imply agreement; for instance if someone measured consistently twice as much as the co-observer, the correlation will be perfect but agreement would be very poor. The most reliable statistical methods for reporting measurement repeatability are those that demonstrate the magnitude of measurement error (within-subject variation) in the context of the natural size variability of the measured structure (between-subjects variation). These can either be numerical methods like the intraclass correlation coefficient[107] and within-subject coefficient of variation[108] or graphical representations like the Bland-Altman plot[109] of mean differences and their 95% limits of agreement.

Another potential pitfall is when measurement repeatability is assessed performing repeat segmentations on the same 3D volume for each case in a study. This can assess error during the segmentation process, but ignores a second source of variation - namely error introduced during the volume acquisition. The latter can be a considerable source of error as a result of fine fetal movement artifact. When repeating the segmentation on the same saved volume, repeatability will appear high, but because the actual volume acquisition is not accounted this is an overestimation of agreement. For instance the intraobserver coefficient of variation for repeat brain volume measurement is only 2.2% when segmenting the same 3D dataset twice[51], which shows strong agreement; however it becomes 10.2% when segmenting two separately acquired 3D volumes of the same fetal head[51]. The appropriate methodology for repeatability studies should therefore involve a minimum of two volume calculations performed on separately acquired volumes; these should be performed by the same observer in order to assess intraobserver error, but also at least one different observer in order to assess interobserver agreement.

Ultimately, every effort should be made that a novel volumetric methodology is validated in vivo. Measurement error, when using automated 3D transducer systems, is more likely to arise from interpretation of the ultrasonic image rather than a technical flaw within the

scanner or its operating software. In vitro validation can only assess the performance of the technical apparatus under optimal experimental conditions. Measures of in vitro accuracy cannot be extrapolated when measuring a specific fetal structure with specific imaging limitations.

Accuracy should be tested in vivo against an independent method of volumetric assessment of the same structure: a gold standard. There are a number of ways to achieve this. One option is to compare ultrasonographic biometry with direct evaluation of fetal structures post mortem, after a planned pregnancy termination. A study of this nature can be challenging to organize. Possible limitations include ethical dilemmas in carrying out research that involves handling of human tissue, the gestational age range in terminations and the fact that these may not be representative examples of healthy on-going pregnancies. Ethical and practical considerations relating to human tissue handling can be overcome to a certain extent with the use of non-invasive imaging modalities, such as fetal MRI or post mortem CT[110]. MRI can achieve good visualization of soft tissue structures, especially of the central nervous system[111]. The lack of fetal radiation risk means that it may also be applied in healthy, on-going pregnancies. It has already been used in order to validate lung volume measurements, both in normal pregnancies and those affected with lung hypoplasia[66, 112]. Conversely, CT cannot be used in a research context on live pregnancies, although it can provide excellent visualization of skeletal structures[113]. However, it may be a suitable non-invasive option for assessing stillborn babies following spontaneous pregnancy loss or termination. A reasonable criticism when using CT or MRI as a gold standard is that these are not necessarily validated methods for fetal volumetry. However, they are 3D modalities with considerably less image artifact and far greater resolution and thus reasonably reflect the true situation; they are certainly a better option than no validation at all.

1.2.4 Conclusions

Studies of fetal organ volumetry are currently subject to considerable methodological heterogeneity. Wide discrepancies in reported normal volumes make diagnosis of pathological organ growth extremely difficult. There is thus a clear need for standardization of volumetric methodology. This involves clear definitions of imaging planes and the anatomical landmarks used for measurement, robust assessment of method repeatability using a recommended experimental design and appropriate statistical tools, and method validation both in vitro and, more importantly, in vivo. Standardization of 3D image file formats and measurement software provided by all ultrasound manufacturers, even if this would not fully eliminate measurement variation, would promote collaborative research, with sharing of stored 3D image data, and would enable comparisons of data obtained using different systems. Such collaboration and comparison would help to establish volumetric methods that are not company specific and would provide a basis for universal standards of quality assurance and training in 3D ultrasound imaging.

1.3 Fetal bone development and the role of vitamin D

During embryogenesis, pluripotent cells are recruited in order to form a structural model of the future skeleton. This is a four-stage process[114] which consists of:

1. *Migration* of primordial cells; those cells originating from the neural crest will form the craniofacial skeleton, whereas cells from the paraxial and lateral plate mesoderm will give rise to the trunk and appendicular skeleton respectively.
2. *Induction* of those primordial mesenchymal cells, through complex signalling interactions with their overlying epithelia.
3. *Condensation* of the mesenchymal cells into a template of the future bone.
4. *Differentiation* of those templates into bone, either via the *endomembranous* or the *endochondral* pathway.

Environmental and nutritional influences during gestation have a different impact on the endomembranous and endochondral bones. Furthermore, ultrasound visualization and the ability to perform measurements are different in those two bone types. An understanding of their basic physiology is therefore necessary in order to appreciate the existing evidence, as well as the aims of this study.

1.3.1 Endomembranous bone differentiation

The flat bones of the cranial vault, clavicle, mandible and parts of the facial skeleton are developed through the endomembranous pathway[114]. The earliest bone template first appears as a mesenchymal condensation; this mesenchymal model is then differentiated directly into bone. This is in contrast with the rest of the endochondral skeleton, where the mesenchymal model is quickly replaced by an intermediate cartilaginous template. Continuing membranous bone growth of the skull during pregnancy and postnatal life is achieved by deposition of newly formed bone along the sutures and fontanelles.

The histological origin of membranous bone during the embryonic period is less well understood than that of endochondral bone. The embryonic progenitors of the orofacial endomembranous bones are cells from the cephalic neural crest[115]. These cells migrate under the inductive effect of the overlying epithelium and then condense to form the mesenchymal template[115, 116]. The origin of the flat bones of the cranial vault is contentious[116]; some research suggests that these may also be derived from the neural crest[117] whereas an alternative theory proposes that they may be of *mesodermal* origin instead[115]. Regardless of their embryological progenitor lineage, those mesenchymal condensations differentiate directly into osteoblasts, which are then responsible for the formation of the calvarial bones.

Sutures and fontanelles, which join the neighbouring skull bones, remain patent and function as a source of continuing skull expansion during pregnancy, but also during postnatal life. Bone formation at the level of sutures is a process tightly regulated by growth and transcription factors[116]. Constant production and degradation of those

signalling molecules results in a concentration gradient between the approximating bone fronts. This gradient maintains a balance between mesenchymal cell *proliferation* (which ensures patency) and *osteogenic differentiation* (which promotes ossification). Skull expansion is thus achieved in order to accommodate the developing brain and also to maintain symmetrical growth of facial features.

Various genetic syndromes have been described, whereby this balance of signalling mechanisms is disturbed[118]. For instance, mutations which lead to loss-of-function of the MSX1 and MSX2 genes[119] are associated with impaired skull ossification. Conversely, MSX2 over-expression[120] (gain-of-function mutations) is linked with premature suture closure. Various craniosynostosis syndromes such as Pfeiffer, Crouzon and Apert syndromes have been associated with FGFR1 and FGFR2 gene mutations[121, 122]. FGFR3 mutations do not appear to affect membranous bone growth, but have a role in endochondral development, as they can cause limb shortening and dwarfism.

1.3.2 Endochondral bone differentiation

The majority of the human skeleton, including the axial skeleton of the trunk and the long bones of the limbs, develop through the process of endochondral ossification[114]. Their embryological progenitor are mesodermal cells: those arising from the *paraxial mesoderm* condense into somites which are precursors of the vertebrae and associated trunk musculature[123]; whereas those located within the *lateral plate* condense into limb buds, which eventually differentiate into long bones and their overlying musculature and dermis[124].

The process of endochondral differentiation is summarized in Figure 4. Each long bone template initially appears as a condensation of mesodermal cells; these cells then differentiate into chondroblasts, which lay down a cartilaginous template rich in type II collagen and matrix proteins. The superficial cell layer enveloping the diaphysis (perichondrium) differentiates into osteoblasts, forming the diaphyseal “bony collar”. Shortly thereafter the *primary ossification centre* appears: the cartilage body of the

diaphysis is invaded by capillaries and the chondrocytes differentiate into osteoblasts; mineral deposition completes the transformation of the cartilage model into bone. For the remainder of the pregnancy a long bone consists of the ossified diaphysis, which appears bright and echogenic on ultrasound; and the cartilaginous epiphyses which are not ultrasonically visible. These only begin to ossify near term, with the ultrasound appearance of the *secondary ossification centres*, after 34 weeks gestation[125].

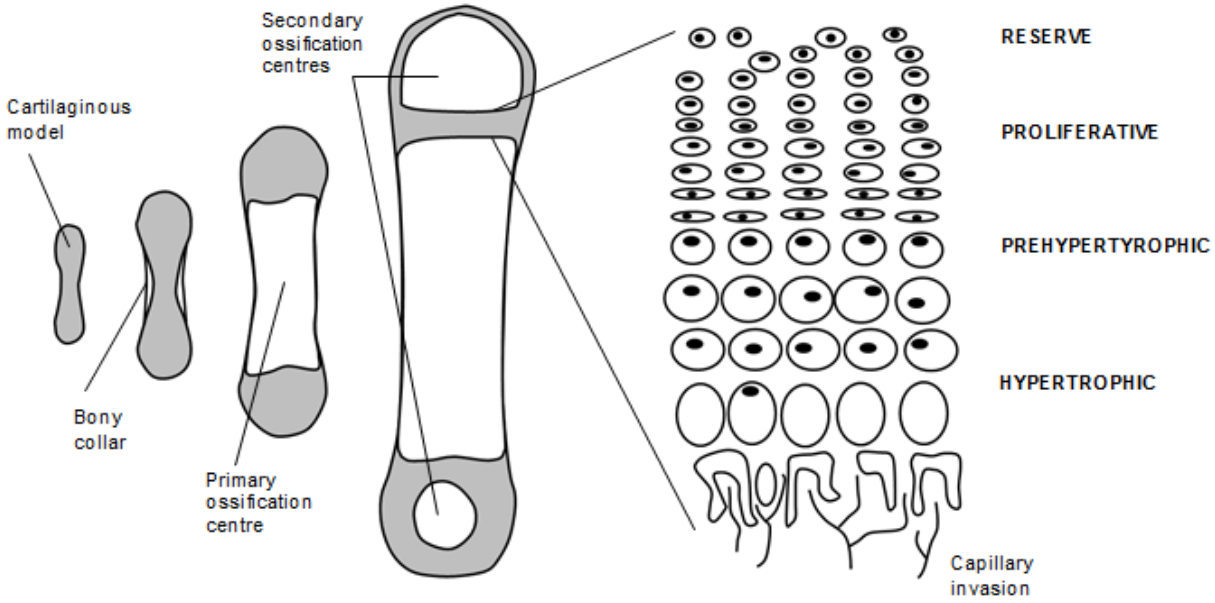


Figure 4 Endochondral ossification and the growth plate

A complex interplay of growth and transcription factors is responsible for the appearance and subsequent development of endochondral bones along the three spatial axes[114, 126]: the proximal-distal, dorsal-ventral and medial-lateral axis. Hox gene transcripts are essential for the cranio-caudal development of the spine and the elongation of the limbs along the proximal-distal axis[114]. Sonic hedgehog gene (Shh) transcripts are key players in the dorso-ventral patterning of midline structures as well as the medio-lateral development of distal limbs[114]. Shh mutations are associated with midline skull anomalies (holoprosencephaly, frontal bossing and cyclopia) and limb defects (short limbs, polydactyly and syndactyly)[114].

1.3.3 The growth plate

Longitudinal bone growth during the prenatal and postnatal period is achieved at the level of the cartilaginous *growth plate,* which is located on either end of the ossified diaphysis (Figure 4). This plate consists of chondrocytes layered into four zones[127]; on the epiphyseal side the *reserve zone* is followed by the *proliferative zone* and they both act as a reservoir for division of cartilage stem cells (chondroblasts) which are responsible for bone elongation. The newly formed chondrocytes then progress through the *pre-hypertrophic* and then the *hypertrophic* zones towards the diaphyseal side, whereby they are committed into progressive ossification.

Ossification is achieved through programmed chondrocyte death (apoptosis), capillary invasion and infiltration by osteoblasts; these lay down the osteoid matrix, which becomes calcified through mineral deposition. Osteoclasts are also present, and these mediate bone remodelling. Long bone growth is thus achieved at the diaphyseal pole of the growth plate through a tightly regulated balance between chondrocyte *proliferation* (which promotes elongation) and *differentiation* (which promotes calcification). This balance is achieved by local signalling mechanisms[114, 127] such as transcripts of the Indian hedgehog gene (Ihh), which block chondrocyte differentiation and increase chondroblast proliferation and bone elongation, via secretion of the PTHrP molecule. Conversely, FGFR3 is a negative regulator of growth and several gain-of-function mutations of the FGFR3 domain can cause genetic skeletal dysplasias, including achondroplasia[128], thanatophoric dysplasia types I and II[129] and hypochondroplasia[130].

Long bone growth at the level of the growth plate is also regulated by well described hormonal pathways. These include the growth hormone, which promotes endochondral growth via IGF1; estrogens, which appear to have a regulatory function, both in females and males, by facilitating ossification and increasing bone age; vitamin D, which is a key player in the provision of mineral as well as in the growth plate biology; and thyroid and steroid hormones which have considerable effects in bone growth when they are

pathologically raised or low[127] although much less is known about their physiological role in healthy individuals.

1.3.4 The role of vitamin D

Vitamin D3 or cholecalciferol is synthesized by human skin cells, through UVB-induced photo conversion of 7-dehydrocholesterol[131]. It can also be obtained through a diet rich in animal and fish fat. Vitamin D2 or ergocalciferol is a similar compound of plant origin, which is at least three times less biologically active compared with D3, when given in supplement form[132].

Whether endogenously synthesized or acquired through diet, D is then transformed into 25(OH)D by the addition of a hydroxyl group at the 25' position[133]. Although 25(OH)D is not biologically active, it represents the body's main store of vitamin D and it is also the main circulating form which is usually bound to a carrier, the vitamin D binding protein (DBP). It can easily be measured in human serum, using widely available biologic assay kits.

The biologically active compound 1,25 $(OH)_2$ D is generated by the addition of a second hydroxyl group at the 1' position, under the action of a renal hydroxylase[134]. The production of 1,25 $(OH)_2$ D is tightly coupled with its inactivation into 24,25 $(OH)_2$ D by a different 24' hydroxylase, forming a regulatory feedback loop.

Bone consists of a *cellular component* (osteocytes, osteoblasts and osteoclasts), an *organic matrix* (type I collagen and other proteins) and the *inorganic mineral component* (mainly calcium and phosphorus in the form of carbonated hydroxyapatite). Vitamin D plays a central role in the regulation of all three components, but primarily ensures a steady provision of mineral for the expanding skeleton. As a fat soluble vitamin, its mode of action is mainly exerted by binding onto the intracellular vitamin D receptor[135] (VDR) following diffusion through the cell membrane. The vitamin D - VDR complex acts as a gene transcription factor[136] which promotes gene expression in several target tissues. The main ensuing effects include the increased provision of mineral via intestinal absorption of

calcium and phosphorus[137] and calcium retention at the kidneys[137]; differentiation, apoptosis and calcification at the level of the hypertrophic zone of the growth plate[127]; activation of osteoblasts, which increases the cellular bone component and of osteoclasts, which is necessary for bone remodelling[137]; finally, non-skeletal effects such as modulation of immune and inflammatory responses[135]. Those combined effects of vitamin D drive long bone growth, through endochondral ossification at the level of the growth plate.

There is no direct evidence to suggest that vitamin D plays any role in endomembranous bone homeostasis. However, clinical studies have demonstrated an association between antenatal vitamin D supplementation in pregnant women and increased ossification of the anterior fontanelle of their babies[138]. In addition, aggressive treatment with vitamin D in X-linked hypophosphatemic rickets has been implicated with premature suture obliteration[139]. These associations indicate a possible effect of vitamin D on suture patency either directly or indirectly as a result of increased mineral provision.

1.3.5 Subclinical vitamin D insufficiency: a modern epidemic

Defining an “optimal” level of serum vitamin D has been the subject of considerable debate. There is little doubt that the clinical syndrome of rickets can manifest when serum levels fall below 25nmol/l. However, it is also well documented that an individual can maintain levels of 100nmol/l or over, by means of adequate sunlight exposure alone[140]. Levels between 25 and 100nmol/l represent a “grey area”, where the threshold level of subclinical vitamin D insufficiency is difficult to establish. In previous studies, deficiency or insufficiency are variously defined using arbitrary cut-offs of <12.5 nmol/l[141], <15 nmol/l[142], <25 nmol/l[143], <27.5 nmol/l[144], <40 nmol/l[145] and <50 nmol/l[146].

An alternative strategy for defining insufficiency is indirectly through the metabolic effects of vitamin D supplementation. Individuals with low levels of vitamin D can have reduced serum calcium and this may lead to raised parathyroid hormone levels (PTH), known as secondary hyperparathyroidism. The therapeutic effect of vitamin D administration in

those individuals can be assessed by measuring the reduction of PTH following supplementation; such physiological studies have shown that there is no added benefit from supplementation when the baseline vitamin D levels exceed 50 nmol/l[147] and this is often considered as the threshold of sufficiency. Levels between 25 and 50nmol/l are therefore regarded as insufficiency, whereas <25nmol/l is frank deficiency[148, 149].

There is huge geographical, cultural and interpersonal variation in natural sunlight exposure, dietary habits and clothing practices. As a result of this variation, the relative importance of dietary acquisition over endogenous vitamin D production is different for different populations. Reduced endogenous production can be caused by decreasing UVB exposure, as in the case of higher latitude, living indoors and clothing habits that limit skin exposure. Dietary acquisition of vitamin D then becomes an increasing necessity in order to maintain optimal bone health. Yet the dietary sources of vitamin D are fairly limited: oily fish, fortified margarines and some breakfast cereals, as well as smaller amounts in red meat and egg yolk[150].

During periods of maximal skeletal growth, especially during infancy, childhood and puberty, deficiency of vitamin D can manifest as rickets, a syndrome characterized by growth failure and the typical appearance is costo-chondral beading, swelling of epiphyses and bowing of weight bearing bones, mainly of the legs. Rickets is now uncommon in the developed world with 2.9 new cases per 100,000 children per year in Canada[151]. However, there are increasing rates of subclinical vitamin D insufficiency, which is not severe enough to manifest as rickets but may still result in suboptimal skeletal mineralization during childhood and puberty[137, 152]. This is of great importance from a public health perspective, as peak bone mass attained during adolescence is a determining factor for the rate of subsequent bone loss and the lifetime risk of osteoporotic fracture. Populations at risk include dark skinned individuals[153] who migrate and live at higher latitude[154]; strict vegetarians[155]; those who systematically avoid direct sunlight exposure[153]; non-supplemented, exclusively breastfed infants[156]; and pregnant women[137], especially those from any of the above risk groups.

1.3.6 Vitamin D insufficiency during pregnancy

The dual role of vitamin D during pregnancy is to maintain maternal skeletal health while at the same time facilitating the mobilization of mineral in order to support the developing fetus. Maternal serum levels of 25(OH)D generally remain stable during gestation in the absence of supplementation[157, 158]. In contrast, the concentration of the biologically active form 1,25$(OH)_2$ D in the mother doubles between the first and third trimesters[158]. Increased availability of the active form drives increased maternal mineral absorption, which in turn supports the fetal calcium accrual, from approximately 50 mg/day at around 20 weeks, to an average of 250 mg/day during the third trimester[159].

It is difficult to estimate the net effect of these changes on the maternal skeleton. Whilst a study of pre- and post-pregnancy bone mineral density measured by dual emission x-ray absorptiometry (DEXA) did not show any significant difference[158], a study of maternal bone density using quantitative calcaneal ultrasound (QUS) demonstrated a slight but significant bone mass reduction during the third trimester[160].

Observational studies of serum vitamin D concentrations during gestation are indicative of high prevalence of subclinical vitamin D insufficiency in developed countries. Dark skinned women living at higher latitudes are at greater risk: 83% of pregnant Pakistani women living in Oslo have a vitamin D less than 30 nmol/l[154]; similarly 80% of veiled or dark-skinned pregnant women at their first antenatal appointment in Melbourne have concentrations less than 22.5 nmol/l[153]. In the context of an interventional study of vitamin D supplementation in South Wales, 50% of a cohort of non-Caucasian pregnant women had levels <20 nmol/l prior to supplementation[161]. These reports are heavily selected as they contain samples of women at the highest risk of antenatal vitamin D deficiency.

Unselected population studies in the US[157], UK[162] and Australia[163] demonstrate an overall prevalence of vitamin D deficiency in pregnancy (<25 nmol/l) between 7% and 18%; vitamin D insufficiency and deficiency combined (<50 nmol/l) range from 33% to 49%. Amongst non-supplemented pregnant women in the US, the rate of insufficiency reaches

66%, compared with 49% in matched, non-pregnant controls[157]. This demonstrates that pregnancy itself puts an added burden on the low vitamin D stores of a significant proportion of women in childbearing age.

It is evident that pregnant women may benefit from vitamin D supplementation, with the objective of optimizing their own skeletal health during pregnancy. Data from interventional[138, 164, 165] as well as observational studies[157] confirm that vitamin D supplementation during pregnancy increases maternal serum and neonatal cord vitamin D concentrations. Adequate supplementation for deficient women is likely to necessitate at least 25 mg (1000 IU) daily and should start ideally in early, rather than late, gestation[161]. The Canadian Paediatric Society now recommends routine supplementation for pregnant and lactating women with 2000 IU of vitamin D daily[166]. The UK National Institute for Clinical Excellence has also endorsed supplementation of high risk pregnant women with 400 IU (10 μg) daily[150]; this recommendation applies to women of South Asian, African, Caribbean or Middle Eastern family origin; those who have limited exposure to sunlight (housebound or covered when outdoors); those whose diet is low in oily fish, eggs, meat, vitamin D-fortified margarine or breakfast cereal; and those with a pre-pregnancy BMI of above 30[150].

1.3.7 Fetal effects of vitamin D insufficiency

Although vitamin D insufficiency is common during pregnancy, there are limited data about its effect on the fetus. Femur length (FL) is the most widely used ultrasonographic marker of fetal bone growth, but no correlation between FL and maternal vitamin D level was found on a cohort of 424 women from Southampton, UK[167]. The same study identified a significant negative correlation between vitamin D and the surface area of a cross-section of the distal femoral metaphysis[167]. This effect of vitamin D deficiency was described as “splaying” of the distal metaphysis; however it is not clear how this affects the overall bone mineral content (BMC) and bone mineral density (BMD) of the fetus. BMC and BMD can be assessed postnatally with DEXA, but this radiological technique

involves a small amount of radiation exposure and has not been widely used in the literature to date. In the absence of DEXA it is reasonable to use birthweight as a surrogate marker, given that there is a strong correlation between birthweight and bone mineral content[168].

There is a well-established seasonal variation of birthweight which could be explained by seasonal changes in sunlight exposure and maternal vitamin D levels. Studies from the northern and southern hemispheres, have demonstrated that population birth weights vary throughout the year, with a seasonal periodicity of approximately 30g. For instance, data from the Queensland perinatal register[169] indicate that babies born in the winter months (June to October) are heavier than those born during the summer (January to May); similarly, babies born in Northern Ireland during the winter months (January to April) are heavier than those born during the summer (June to September)[170]. It can be inferred that sunlight exposure in early pregnancy - rather than at the time of birth - replenishes the maternal vitamin D stores and therefore drives fetal skeletal growth during the remainder of the pregnancy.

A survey of 971 pregnant women in Australia[163] demonstrated such a seasonal periodicity of serum vitamin D levels; higher mean concentrations were noted in the sunny months of January to May. Babies born to vitamin D deficient mothers also had lower birthweight, with an adjusted mean difference of 151 g. Finally, a UK cohort of 198 mother – baby pairs were followed up postnatally until the age of 9, where a DEXA of the child was performed; it was demonstrated that maternal vitamin D insufficiency during pregnancy was associated with reduced BMC in childhood[162].

The above evidence demonstrates that naturally occurring variation of maternal vitamin D levels during pregnancy has a measurable impact on anthropometric measures at birth and that this effect is still strikingly demonstrable into childhood age. This is an important public health issue. A significant body of epidemiological data supports the theory of fetal programming according to which suboptimal development of the fetal skeleton may increase the lifetime risk of osteoporosis[171]. Intrauterine life - like adolescence and puberty

- is a period of intensive skeletal expansion. Failure of the developing fetus to achieve its potential for bone growth can re-program its developmental trajectory. This is similar to an adolescent failing to achieve optimal peak bone mass following puberty. Such an individual may be predisposed to a higher risk of fracture in later life, once this trajectory eventually crosses the threshold of clinical osteoporosis.

The evidence supporting the theory of fetal programming comes from several studies linking parameters of fetal growth with infantile weight, bone content in adulthood and eventually the incidence of osteoporotic fracture in later life[171]. The first link in this chain of evidence is the observation that BMC at birth is positively correlated with parental birth weight and maternal adiposity[168] and negatively correlated with maternal smoking and strenuous exercise during pregnancy[168]. Infant and childhood growth are in turn linked to bone mineral content in adult age: observational cohort studies have demonstrated associations between weight in infancy and the femoral and spine BMC in adulthood[172, 173]. Finally, a linkage study of childhood growth records from the 1920s with hospital discharge codes for hip fracture seventy years later demonstrated an elevated fracture risk amongst individuals with slow rate of childhood growth[174]. Putting this evidence together, it is plausible that the suboptimal developmental trajectory which eventually leads to the clinical manifestations of osteoporosis may start as early as the intrauterine period.

1.3.8 Vitamin D supplementation in pregnancy: fetal effects

It is not clear whether vitamin D supplementation in pregnancy has any impact on markers of fetal skeletal growth, such as birthweight or neonatal BMC. There is some animal experimental data to support this hypothesis: administration of vitamin D to female pregnant rats was shown to increase the dry tibial weight, tibial ash weight and the whole body weight of the 28-day old offspring[175]. However the evidence in humans is less clear. A single, non-randomized, cohort study assessed the neonatal BMC between supplemented and non-supplemented pregnant women; an unvalidated method of

radioisotope absorptiometry of the baby's forearm was used but this did not demonstrate any difference between the study groups[176].

We performed a literature review for studies of maternal vitamin D supplementation during pregnancy using birthweight as an outcome measure. We identified six studies[138, 164, 165, 176-178] and one Cochrane systematic review[179]. One of the primary studies reported no significant difference in birthweight but did not provide the birthweight data[164]. We performed a meta-analysis of birthweight data of all remaining five studies: this indicated a significant increase of birthweight by supplementation, with a mean difference of 138 g (Forest plot shown in Figure 5).

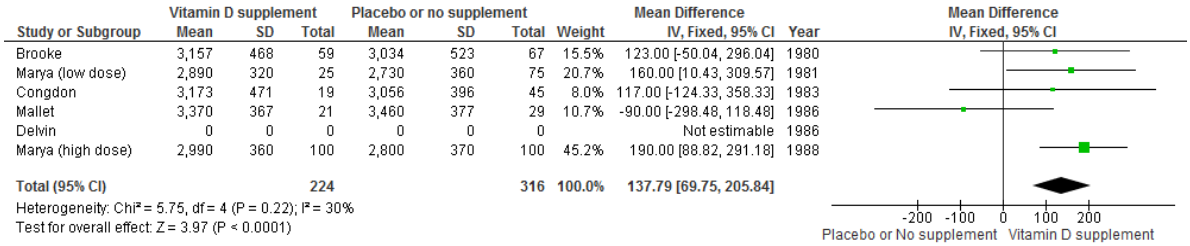


Figure 5 Meta analysis of trials of Vitamin D supplementation in pregnancy, using birthweight as the outcome measure: all studies included

Only two of the included studies are well-described RCTs[138, 165] whereas the other three are cohort or interventional studies[176-178] with high risk of bias. Separate analysis of the two RCTs, as performed by the Cochrane systematic review, does not show any significant difference in birth weights between babies of supplemented mothers and those receiving placebo (Figure 6): one study found a non-significant trend for reduced birth weight in the supplementation group[165] and one study showed a trend for increased birth weight[138]. In the treatment arm of the latter study there was a significant reduction of the anterior fontanelle surface area measured clinically at birth; and a non-significant reduction of small for gestational age babies from 28.6% to 15.3%. Moreover, babies from supplemented pregnancies were significantly heavier at the 1 year postnatal follow up[180].

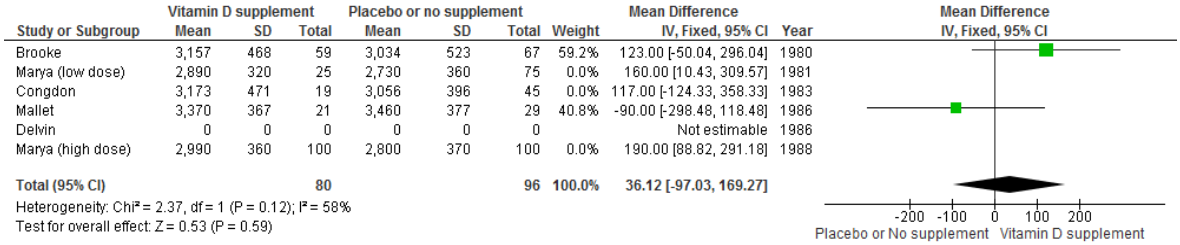


Figure 6 Meta analysis of trials of Vitamin D supplementation in pregnancy, using birthweight as the outcome measure: only studies at low risk of bias included

In summary, there is reasonable evidence to suggest that maternal vitamin D deficiency is linked with reduced birth weight and BMC of the offspring. However it has not yet been established whether vitamin D supplementation during pregnancy improves any of those outcomes.

1.4 Background summary, hypotheses and aims

Different developmental pathways exist for fetal skeletal differentiation. Maternal vitamin D plays a central role in the physiology of endochondral bone, but some evidence suggests it may also affect the development of endomembranous bone. 3D ultrasound techniques have been employed in the study of both types of bone. A typical example of endochondral bone is the femur, yet the femur length does not have any significant correlation with maternal vitamin D levels. A potential alternative is the femur volume, but the existing methodology for measurement in 3D has shown poor interobserver agreement.

There is a single study that suggests there may be a link between vitamin D status and the fontanelle in the neonate; this raises the possibility that fetal fontanelle area could represent a marker of membranous ossification. The anterior fontanelle has previously been visualized prenatally with rendered 3D ultrasound, but not in relation to vitamin D status. However the ability to visualize the anterior fontanelle is often hampered by cephalic fetal presentation. The sphenoidal fontanelle is easier to visualize, but its surface

area has not been studied to date; in addition, fontanelle area measurement using rendered 3D ultrasound is a method which has not been validated.

1.4.1 Hypotheses

- Fetal imaging with 3D ultrasound is more sensitive than 2D, in assessing fetal skeletal growth and mineralization.
- Femur volume and sphenoidal fontanelle surface area are useful markers of fetal skeletal growth.
- There is a correlation between those markers and maternal serum vitamin D concentration.

1.4.2 Aims of the thesis

- To develop reliable methodology for fetal femur volume and sphenoidal fontanelle area measurements using 3D ultrasound.
- To create a reference range of normal femur volume and sphenoidal fontanelle area in healthy, uncomplicated pregnancies.
- To assess femur volume and sphenoidal fontanelle area in correlation with maternal vitamin D concentration and other maternal predictors of bone size, in a cohort of pregnant women.

1.5 Publications and contributions

The author performed the literature review and created all tables and figures included in this chapter. Material from this chapter was included in the following publication:

- Ioannou C, Sarris I, Salomon LJ and Papageorghiou AT. A review of fetal volumetry: the need for standardization and definitions in measurement methodology. Ultrasound Obstet Gynecol. 2011 Dec;38(6):613-9. doi: 10.1002/uog.9074

2 Sphenoidal fontanelle surface area

This chapter describes the principles, method development and results of sphenoidal fontanelle area measurement using 3D ultrasound with maximum rendering.

2.1 In vitro validation of surface area measurement using a phantom

2.1.1 Introduction

Fetal biometry usually relies on 2D imaging planes of fetal organs, amenable to linear measurements only (length or circumference). 3D ultrasound permits the visualisation of complex structures or entire organs, whereby volume or surface area measurements may be performed. One particular area of interest is in the assessment of the fetal skull, cranial sutures and fontanelles[34-36, 40]. Most fontanelles and sutures can be reliably demonstrated using rendering techniques throughout the second half of gestation[35]. Visualization of fontanelles on rendered 3D images may be affected negatively by advancing gestational age and cephalic presentation, due to increasing difficulty of scan acquisition of the entire fetal head[35]. A pocket of amniotic fluid between the transducer and the fetal skull often optimizes the rendered image and therefore reduced amniotic fluid volume - a frequent occurrence with advancing gestation - is also likely to affect the visibility of these structures[35].

Surface area measurements of the anterior fontanelle, using rendered 3D ultrasound, have been described and relevant normograms have been published[40]. Potential clinical applications include screening for Downs syndrome, as these fetuses have larger fontanelles[41]; or in the diagnosis of various craniosynostosis syndromes[38, 39]. Fetal fontanelle size and development may also be influenced by maternal vitamin D status[138].

However, the use of 3D ultrasound has not been validated for fetal biometry on *rendered* images.

While volume calculation using 3D ultrasound has high reproducibility and validity[42], surface area measurements when using rendered images cannot be assumed to have the same accuracy. Rendered scan images are reconstructed pictures, whereby several points of ultrasound reflected from different depths are projected onto a single imaging plane, and this has the potential to lead to significant error.

In general, sources of measurement variation may include error during calliper placement and variability during generation of the scan image; for example due to an incorrect plane or the effect of fetal breathing on abdominal circumference. For rendered 3D images in particular, measurement error may be the result of fetal movement artifact during volume acquisition. There is also the possibility of true biologic variation; thus, the head diameters and fontanelle size may vary as a result of moulding and external pressure because the calvarial bones and fontanelles are compressible structures.

The aim of this study was to assess the repeatability and validity of surface area calculation on rendered 3D images, using an in vitro phantom that simulates skull fontanelles.

2.1.2 Methods

This was an in vitro phantom validation study. A commercially available ultrasound scan machine with a mechanical 3D transducer was used (Philips HD-9 with V7-3 transducer, Philips Ultrasound, Bothell, Washington, USA). A glass test-tank containing a solution of 7% glycerol in water was the scanning medium [42, 181]. The phantom consisted of a flat tile made of rigid vinyl plastic (1 mm thickness) to simulate skull bone (Figure 7). Four holes of standard geometrical shapes were cut out on the tile representing fontanelles.

The shapes chosen were a regular pentagon, trapezoid, isosceles triangle and a parallelogram. The phantom was positioned horizontally into the test-tank and time was allowed for the solution to de-gas. The ultrasound transducer was then immersed 2 cm

into the solution and a volume scan for every shape was obtained with a vertical angle of insonation. In order to assess whether the direction of the sweep has an effect on measurement, the 3D acquisition was repeated with the transducer rotated by 90^{o} (Figure 7, x and y axes). This process was repeated with the phantom positioned at four different depths: 10 cm, 8 cm, 6 cm and 4 cm from the transducer. A second sonographer scanned the phantom independently and in a blinded fashion, using the above protocol.



Figure 7 A flat vinyl tile with four holes of perfect geometrical shapes was scanned in a test tank by two sonographers, in four different depths and using two sweep directions: along the x axis and then along the y axis

Each scan was obtained using a varying sweep angle in order to acquire one shape per image and a default sweep speed of 5 seconds for a 65^{o} sweep. A low gain setting was used during scanning to minimize ultrasound artifact due to reflexion from the tank walls. The 3D image was displayed using maximum rendering mode. Brightness and contrast were adjusted in order to enhance the shape demarcation: image contrast was set at maximum and image brightness was kept at an average setting. Each rendered image was extracted in a digital format (jpeg file) and analysed on a PC using image viewing software for measurement (Escape Medical Viewer, version 3.2.3, www.escape.gr). Every image was calibrated for measurement using the scale displayed on the left side of the scan picture. Manual tracing of the phantom fontanelle was then performed using the polygon measuring tool (Figure 8).

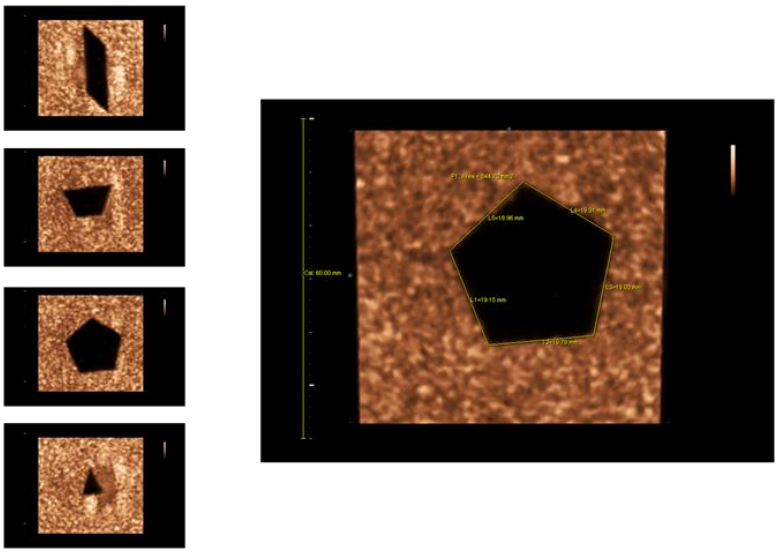


Figure 8 Appearance and measurement of the phantom on rendered 3D ultrasound

Images were processed by two operators independently of each other, blinded to each other's results and also to the true size of the fontanelles. In order to assess intraobserver and interobserver repeatability, all images were traced twice by operator 1 and once by operator 2. In order to assess validity, scan measurements were compared to the true surface areas; the shape dimensions (sides, base and height) were measured by both observers directly onto the phantom and the true surface areas were calculated using standard geometric equations as shown in Table 3.

Table 3 Shapes and their true surface areas calculated using standard mathematical equations; S = surface area, t = side length, h = height, b = base

Shape	Equation	Surface area (mm^2)
Triangle (isosceles)	$S = \frac{1}{2}bh$	95.3
Trapezoid	$S = \frac{1}{2}h(b_1 + b_2)$	325.5
Parallelogram	$S = bh$	471.9
Pentagon (regular)	$S = \frac{t^2\sqrt{25+10\sqrt{5}}}{4}$	654.2

Data were analysed with PASW Statistics 18.0 (SPSS Inc. Chicago, Illinois 60606, USA). Intra and interobserver measurement repeatability were expressed as within-subject coefficients of variation (CV)[108] and intraclass correlation coefficient (ICC) with 95% confidence intervals (CI)[107]. Interobserver percentage differences were also plotted using the method described by Bland and Altman[109]. Validity was expressed as the percentage difference of the measured from the true surface area. Percentage differences at different scan conditions (depth, shape, sonographer, sweep direction) were assessed for normality of distribution using Kolmogorov-Smirnov and Shapiro-Wilk tests. As data were not normally distributed, they were compared using non-parametric significance tests: Kruskal Wallis test for comparisons between multiple groups (depth and shape effect) and Mann Whitney U test for paired comparisons (sonographer and sweep direction). Mann Whitney U was also used post-hoc, to test all six possible paired comparisons of four depths and of four shapes. In the latter case, significance was set at $p = 0.05/6 = 0.008$ using the simple Bonferroni correction method[182].

2.1.3 Results

There were a total of 64 volume scans; 4 shapes scanned at 4 depths, by 2 sonographers using 2 transducer orientations. The true surface areas of the four phantom fontanelles are listed in Table 3. Intra and inter observer repeatability measures are shown in Table 4 and were all satisfactory. Coefficients of variation ranged from 1.6% to 5.5% and intraclass correlation coefficients were in excess of 0.99.

Bland-Altman plots of the interobserver percentage differences showed that the 95% limits of agreement between two observers were 6.2% to -14.9% when measuring the same scan image; while for measuring an independently acquired scan image the limits of agreement were 16.2% to -17.5% (Figure 9).

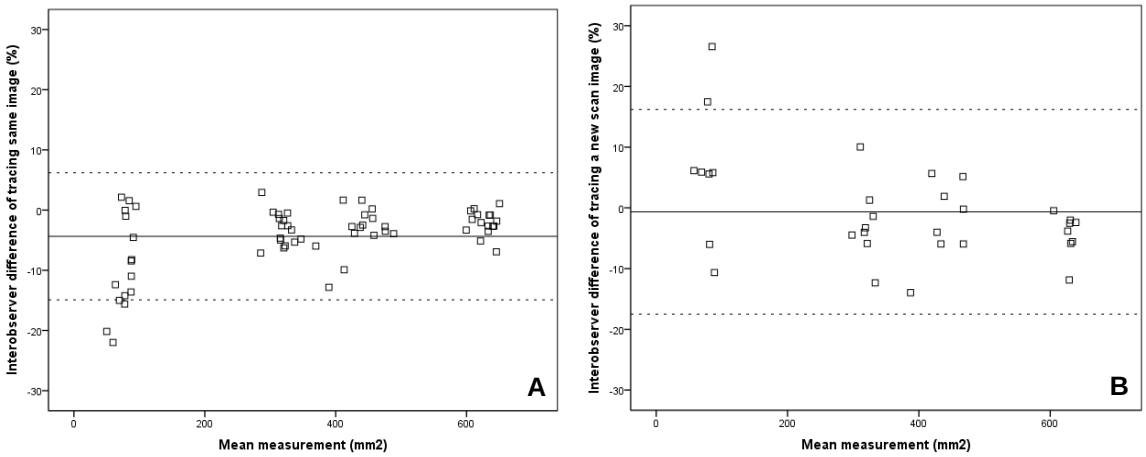


Figure 9 Bland-Altman plots of interobserver percentage differences when tracing the same scan image (A) and on a newly acquired image (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

Table 4 Repeatability of surface area measurement in vitro; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval

		Tracing same scan image	Tracing new scan image
Intraobserver	**CV**	1.6%	5.5%
	ICC	0.999	0.992
	(95% CI)	(0.999 – 0.999)	(0.974 – 0.999)
Interobserver	**CV**	3.1%	5.0%
	ICC	0.997	0.992
	(95% CI)	(0.975 – 0.999)	(0.981 – 0.996)

In order to assess validity, the first ultrasound measurement by the first observer was compared to the true surface area. The median percentage difference from the true was -3.5%, mean percentage difference -5.9% and the 95% limits of agreement were -23.9% to 12.1%. The largest percentage difference however was noted for the triangle, which was the smallest shape tested: median difference for the triangle was -12.1% vs. -0.6% for the trapezoid, ($p < 0.0001$) and -2.4% for the pentagon, ($p = 0.002$) (Figure 10).

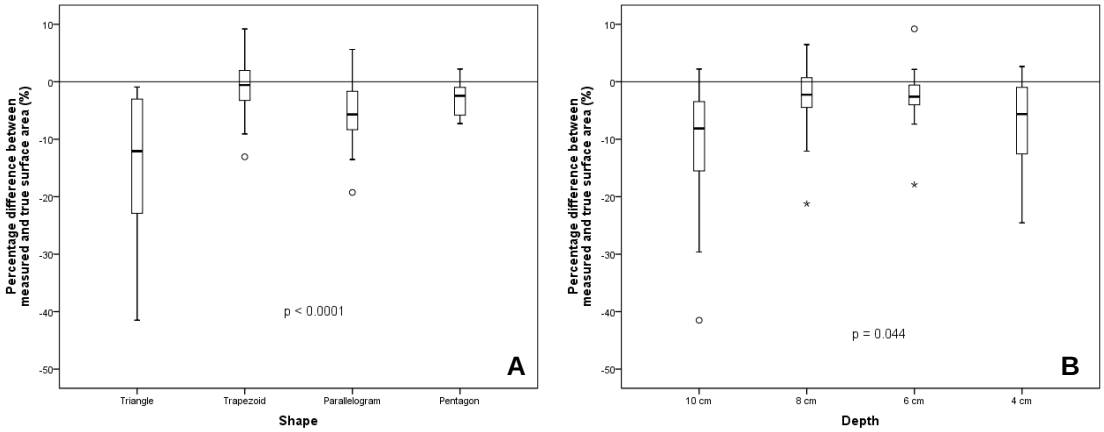


Figure 10 Validity expressed as the percentage difference between the measured and true surface area versus shape (A) and depth (B); box plots demonstrate median, interquartile range, outliers (o) and extreme values (∗); p value calculated using the Kruskal Wallis test

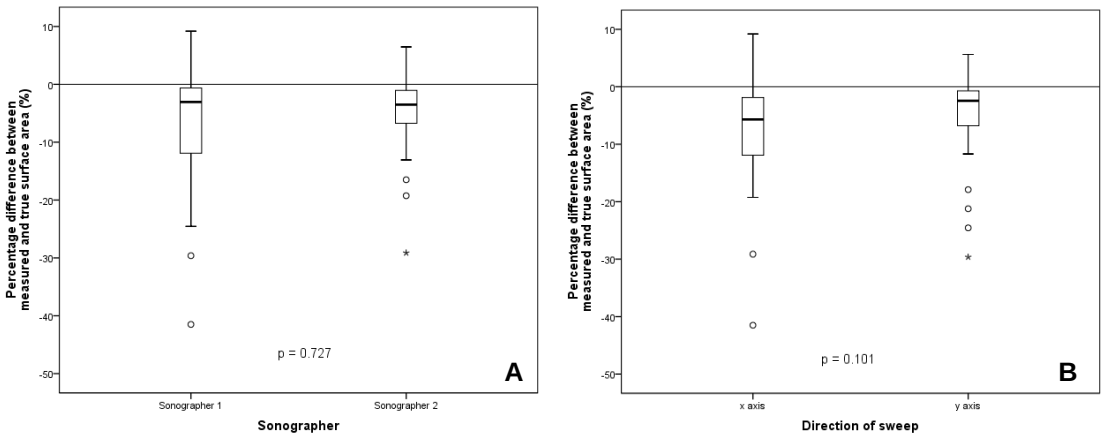


Figure 11 Validity expressed as the percentage difference between the measured and true surface area versus sonographer (A) and direction of 3D sweep (B); box plots demonstrate median, interquartile range, outliers (o) and extreme values (∗); p value calculated using the Mann Whitney U test

Depth also affected measurement accuracy to some extent (p = 0.044). The highest measurement error (median -8.1%) was noted when the phantom was furthest away (10 cm) from the transducer, though none of the pair wise post hoc comparisons reached

Bonferroni-corrected statistical significance (Figure 10). The smallest median percentage differences were noted for intermediate depths (-2.2% at 8 cm and –2.6% at 6 cm). Neither the operator performing the scan (p = 0.727) nor the direction of 3D sweep (p = 0.101) had any appreciable effect to the validity of the technique (Figure 11).

2.1.4 Discussion

This study demonstrates that surface measurement on rendered 3D ultrasound images is accurate and reproducible in vitro. To the best of our knowledge this is the first attempt to validate surface area measurements on rendered scan images.

Fontanelles are not entirely flat structures. Even though they can be demonstrated in part with 2D ultrasound, scanning on a single plane may not achieve adequate and consistent visualization. Using rendered 3D ultrasound instead, successful visualization is feasible in 82-100% of cases[35]. Fetal anterior fontanelle measurement using this technique has been described previously and the ICC for intra- and interobserver repeatability were 0.87 and 0.83 respectively[41]. These are lower than ICC for *volume* calculation, quoted in the literature as being in excess of 0.997 in vitro[42] and in vivo[183, 184]. Volume calculations are achieved using measurements performed on the constituent 2D planes of a 3D scan; conversely fontanelle area measurements are taken on the rendered image. The different ICC figures therefore support the hypothesis that those two different methodologies are subject to different measurement bias.

Our study makes a clear distinction between two sources of measurement variation: as a result of the manual tracing of the scan image and following measurement of a newly generated image. Repeatability of tracing the same image is overall excellent both within and between observers. Variability of tracing a newly acquired scan image was considerably greater. The measurement variation introduced during the scan image generation is most likely due to movement artifact which is an inherent limitation of 3D ultrasound. A 3D volume acquisition usually takes 2-5 seconds, depending on scan machine settings (sweep angle and speed). Any fetal or transducer movement during

acquisition may result in image distortion, which may affect measurement reproducibility. In this experiment there were no “fetal” movements. However movement artifact may still be caused by the sonographer holding the transducer while obtaining the 3D volume.

This study shows that there is a systematic underestimation of the true surface area, by an average of -5.9%. This compares reasonably to the percentage error between +1.4% and +4.1% for *volume* measurement using 3D ultrasound published elsewhere[42]. There is also a degree of random error as evidenced by wide limits of agreement. Despite these figures for random and systematic error, this technique generated a satisfactory figure for intraclass correlation coefficient at 0.992. This means that within-subject measurement bias was very small when compared to the size variation between subjects in our experiment. The phantom shapes were purposely designed so that the range of surface areas (95.3 - 654.2 mm^2) matched the reported values of the anterior fontanelle surface area in vivo[40]. It is interesting that reduced accuracy was noted for the smallest fontanelle (less than 100 mm^2). There was no difference in accuracy between different sonographers or when using a different direction of sweep. The effect of depth was difficult to evaluate: while, overall, there seemed to be a significant difference between groups, pair wise comparison showed a non-significant trend for reduced accuracy at the greatest depth, and this would be logical. However, depth is generally dictated by patient characteristics and this may not be a correctable factor in normal practice; where feasible it would be preferable to avoid surface area measurements at great distances from the transducer, in keeping with general principles of ultrasound.

There are some limitations to this study. Firstly it was carried out using only one ultrasound scanner and mechanical transducer, as specified in the methods section. However there is not a standard digital format for saving, analysing and reporting 3D ultrasound data; image encoding and rendering algorithms vary amongst ultrasound manufacturers. It is therefore possible that different ultrasound equipment may demonstrate different measurement accuracy. We also explored the effect on accuracy of the sonographer, depth, sweep direction and phantom shape using a univariate model. It

is possible that a mixed interaction exists among those four factors; however the experiment was not designed to explore a four-factor interaction model. Finally, this experiment cannot account for any temporal size variation of the surface area of interest, as a source of bias. The phantoms were made of non-compressible vinyl material. It can therefore be assumed that the surface area of each hole was constant throughout the experiment. It is possible however that the fontanelle size may vary in vivo due to moulding of skull bones, fetal position or external abdominal pressure. It is also possible that the demarcation between skull bones and the fontanelle is less clear in vivo and this would make tracing less reproducible. These effects can only be investigated by performing an in vivo reproducibility study, acquiring and measuring multiple scan images for each fetus.

2.1.5 Conclusions

In conclusion, this study demonstrates that surface area measurement in vitro, using rendered 3D ultrasound of a phantom simulating fetal fontanelles, is subject to a low measurement error. This technique is therefore reliable enough to be applied in vivo.

2.2 Sphenoidal fontanelle measurement is not reliable in vivo

2.2.1 Introduction

In the previous study we described and validated in vitro a technique for surface area measurement on rendered 3D ultrasound images[185]. Quantitative assessment of fontanelle size is not routine practice. However there is evidence that fontanelle ossification depends on maternal vitamin D levels[138] and size may therefore be useful as a marker of fetal skull mineralization. The anterior fetal fontanelle area has been assessed during pregnancy and a normogram has been published[40]. However such measurement

necessitates a sagittal head volume acquisition from the top of the skull which is practically difficult, especially in late gestation with cephalic presentation.

A possible alternative is to visualise the sphenoidal fontanelle, because this can be seen using a transverse volume acquisition of the head - in the same orientation as used for measurement of the biparietal diameter. Quantitative assessment of the sphenoidal fontanelle surface area has not been reported so far and its clinical usefulness is uncertain. The objective of this study was to explore the feasibility and repeatability of measurement of sphenoidal fontanelle area throughout gestation using 3D ultrasound with maximum rendering.

2.2.2 Methods

Women with uncomplicated singleton gestation were offered to take part in this study, with informed consent and following regional ethics committee approval. Serial ultrasound scans were performed starting from 14 weeks gestation, approximately every 5 weeks till delivery. A commercially available ultrasound scan machine with a mechanical 3D transducer was used (Philips HD-9 with V7-3 transducer, Philips Ultrasound, Bothell, Washington, USA). During each examination, two transverse head volumes were obtained at the level of the biparietal diameter, which was defined as the plane containing the cavum septum pellucidum one third of the way along the horizontally orientated falx cerebri; visible thalami; and absent cerebellum (Figure 12).

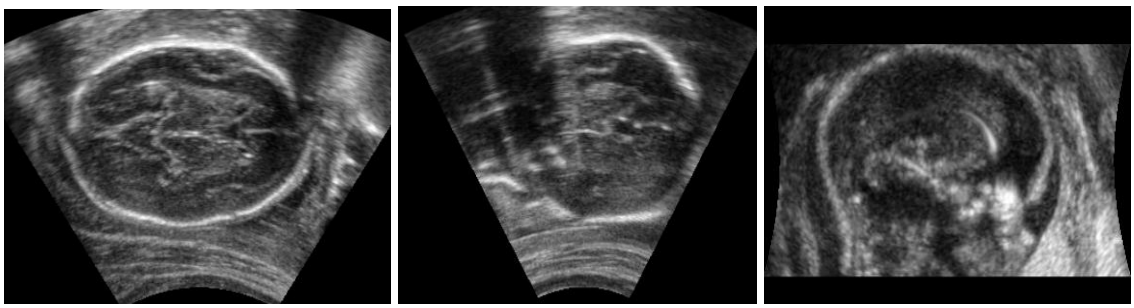


Figure 12 A transverse head volume acquired at the level of BPD is ideal for visualization of the sphenoidal fontanelle

Magnification was set so that the head in 2D fills at least 30% of the screen and a varying sweep angle was used in order to acquire the entire head. Acquisitions were performed during fetal quiet periods with the fetus immobile; a minimum pocket of amniotic fluid between the fontanelle and the uterine wall was allowed where possible. All scans were performed by two obstetricians with experience in 3D ultrasound.

The proximal skull plate was processed using maximum rendering mode and adjacent structures were sculpted away. Brightness and contrast were adjusted in order to enhance the suture demarcation: image contrast was set at maximum and image brightness was kept at an average setting. Each rendered image was downloaded on a PC and analysed using software for biomedical image measurement (Escape Medical Viewer, version 3.2.3, www.escape.gr). The image was first calibrated for measurement using the scale displayed on the left side of the scan picture. Manual tracing of the sphenoidal fontanelle was performed using the polygon measuring tool as shown in Figure 13. All measurements were performed by one operator.

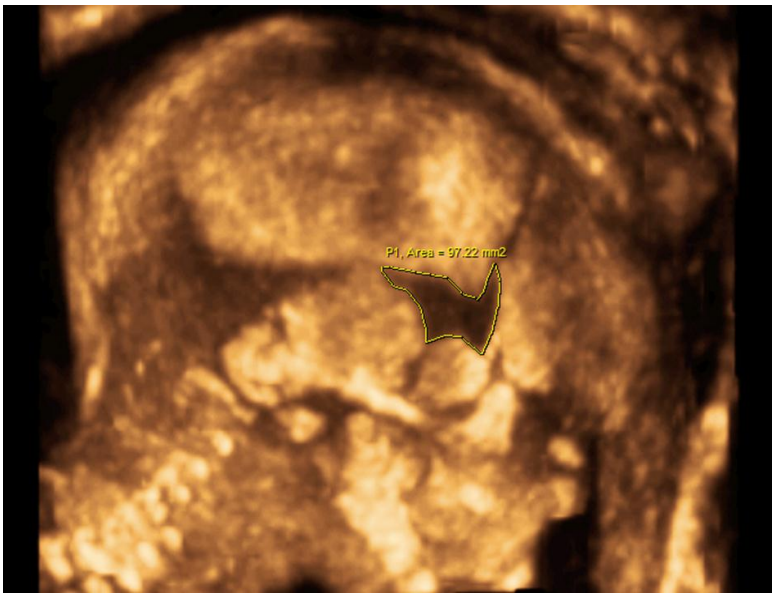


Figure 13 Appearance and measurement of the sphenoidal fontanelle area on rendered 3D ultrasound

In order to assess feasibility, the gestational ages (GA) were divided into consecutive four week blocks, i.e. 14^{+0} - 17^{+6}, 18^{+0} - 21^{+6}, 22^{+0} - 25^{+6} weeks etc. and for every block the absolute and percentage numbers of successful acquisitions were presented. An acquisition was deemed successful if a satisfactory BPD plane was feasible and if the rendered image allowed clear identification of the sphenoidal fontanelle. Two types of measurement repeatability were assessed as in the previous study: repeatability of re-measuring the same scan image and also that of measurement on a newly acquired head volume of the same fetus. There was an interval of a week between first and second measurements in order to avoid recall bias. Repeatability was expressed as the within-subject coefficient of variation (CV) and intraclass correlation coefficient (ICC) with 95% confidence intervals (CI). Percentage measurement differences were presented in Bland-Altman plots. Data analysis and figures were generated with SPSS version 19 (IBM, SPSS Statistics, US).

2.2.3 Results

A total of 141 ultrasound scans were performed in 30 women. Table 5 demonstrates the feasibility of sphenoidal fontanelle measurement within each GA block. Feasibility was highest at 18 – 26 weeks, whereas this dropped after 30 weeks gestation; there was no successful visualization beyond 34 weeks gestation.

Table 5 Feasibility of sphenoidal fontanelle area measurement at different gestational ages; GA = gestational age, n= number of cases

GA	Total cases (n)	Successful cases % (n)
14 – 18	24	83.3% (20)
18 – 22	15	93.3% (14)
22 – 26	21	95.2% (20)
26 – 30	26	69.2% (18)
30 – 34	24	12.5% (3)
34 - 38	20	0% (0)
38 – 42	11	0% (0)

Intraobserver repeatability was assessed on the first 44 consecutive, successfully visualized cases. Mean gestation was 22 weeks, range 16 – 29 weeks. The median fontanelle surface area was 78.9 mm^2. Repeat measurement of the same image had a CV of 17.5% and ICC of 0.883 (95% confidence interval 0.797 – 0.935). However, when the second measurement was performed on an image from a separately acquired volume, the CV increased to 30.7% and the ICC dropped to 0.649 (95% CI 0.441 – 0.792) – shown in Table 6.

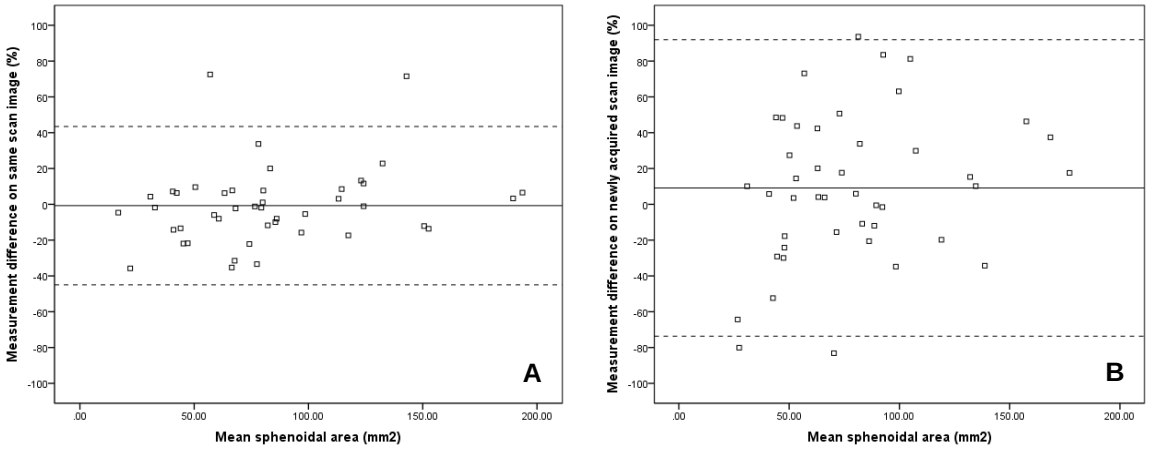


Figure 14 Bland-Altman plots of intraobserver percentage differences when measuring twice the same image (A) and when re-measuring an image from a newly acquired volume (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

Table 6 Intraobserver repeatability of sphenoidal fontanelle area measurement in vivo; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval

		Tracing same scan image	Tracing new scan image
Intraobserver	**CV**	17.5%	30.7%
	ICC	0.883	0.649
	(95% CI)	(0.797 – 0.935)	(0.441 – 0.792)

Figure 14 shows the Bland-Altman plots of percentage measurement differences: when tracing the same image, mean percentage difference was -0.8% (95% limits of agreement -45% to 43.4%); for measurement of a newly acquired scan image, mean percentage difference 9.1% (-73.7% to 91.9%).

2.2.4 Discussion

This study has demonstrated that sphenoidal fontanelle area measurement on 3D ultrasound is possible, but that this is only feasible at less than 34 weeks gestation, and that the repeatability of such measurements is poor. It is the first study to assess feasibility and repeatability of this type of measurement.

Quantitative assessment of fontanelle size could have a number of potential applications such as better characterization of craniosynostosis, Down's syndrome screening and as a marker of fetal bone mineralization. Our previous in vitro validation of this technique using phantom fontanelles showed satisfactory repeatability with CV 5.5%, ICC = 0.992 and satisfactory accuracy with mean percentage difference from the true -5.9%[185]. A previous report on the anterior fontanelle surface area[41] showed acceptable repeatability with intraobserver ICC of 0.87 and mean area 309 mm^2; however repeatability in the latter study was assessed by repeat measurement of the same volume; in addition the authors acknowledged the inability to acquire a satisfactory head volume in a significant proportion of cases. This is because the anterior fontanelle can only be visualized from a sagittal volume taken at the top of the skull and this is usually impossible if the fetus is in a cephalic presentation. Conversely, visualization of the BPD plane from the side of the skull is almost universally possible[35] and this 3D acquisition makes the sphenoidal fontanelle an appealing alternative.

There are two steps involved in measuring the sphenoidal fontanelle. Firstly, a BPD and related volume needs to be obtained; second the rendering needs to allow the sphenoidal fontanelle to be visualised reasonably circumscribed and amenable to measurement. In

our study it was possible to acquire a BPD and the related volume in all but one case. The feasibility of the rendering step was GA dependent: at gestations below 30 weeks, rendered views of the sphenoidal fontanelle were usually successful making measurement feasible. However, with advancing gestation the ability to obtain satisfactory views decreased, and no fontanelle measurements were possible after 34 weeks. Feasibility reported here is lower than in a previous study which quoted excellent visibility beyond 30 weeks[35]. However our definition of feasibility was a stricter one, including not only the visualization but also high image quality which permits measurement.

Not only was feasibility of visualising the sphenoidal fontanelle limited, but measurement repeatability was also poor. This was especially evident when re-measuring the fontanelle on a newly acquired volume of the same fetus. Many previous studies overestimate repeatability when duplicate tracing of the same scan image is reported, but the repeat acquisition of the volume - an important source of error - remains unaccounted for. This is evident from our study: repeat measurement of the same scan image had an ICC of 0.883, and this was very similar to ICC of measurement of the anterior fontanelle reported in a previous study[41]. However, we found that the re-measurement from a newly generated volume resulted in a much lower ICC of 0.649.

Possible explanations for the poor reproducibility include the irregular shape of the sphenoidal fontanelle (Figure 13); and the fact that there is often a lack of clear demarcation between the bone and fontanelle. It was noted that the appearance and size of the fontanelle varied considerably on different acquisitions even if these were done seconds apart, and it is possible that this is due to movement artifact. Finally, the median fontanelle size was around 80 mm^2, and measurement accuracy and repeatability are significantly reduced[185] when the surface area is small (i.e. below 100 mm^2).

This study has some weaknesses. Only one type of ultrasound machine was used and it is possible that different models may produce clearer fontanelle images with superior measurement reliability; however, the machine used is a recently developed and released model and it is unlikely to have led to a significant disadvantage. In addition, we did not

assess interobserver agreement; however, given the fact that intraobserver agreement was shown to be so poor and as interobserver agreement is almost always worse than intraobserver, it was deemed unnecessary to perform such an experiment.

2.2.5 Conclusions

Whilst it is feasible to visualize the sphenoidal fontanelle on rendered 3D before 34 weeks, its surface area measurement is not reliable for clinical use. There is potential to overestimate repeatability when re-measuring the same scan image. Future studies using similar methodology for assessment of fontanelles should provide robust assessment of feasibility and repeatability.

2.3 Publications and contributions

The author conceived and implemented the experiments included in this chapter. All ultrasound scans and surface area measurements were performed by the author and repeated by a second sonographer (Dr I Sarris) for interobserver agreement analysis. Statistical analyses, tables and figures were performed by the author. Material from this chapter was included in the following publications:

- Ioannou C, Sarris I, Yaqub MK, Noble JA, Javaid MK and Papageorghiou AT. Surface area measurement using rendered three-dimensional ultrasound imaging: an in-vitro phantom study. Ultrasound Obstet Gynecol. 2011 Oct;38(4):445-9. doi: 10.1002/uog.8984
- Ioannou C, Sarris I, Javaid MK and Papageorghiou AT. Sphenoidal fontanelle area measurement on rendered three-dimensional ultrasound. Prenat Diagn. 2012 Jun;32(6):592-5. doi: 10.1002/pd.3867

Material from this chapter was also presented by the author under the title “Surface area measurement using rendered three-dimensional ultrasound imaging: an in vitro phantom study” at the 20th ISUOG World Congress 2010, Prague, Czech Republic (oral

presentation) and at the British International Congress of Obstetrics & Gynaecology 2010, Belfast, UK (oral poster presentation).

3 Femur volume measurement using multiplanar segmentation

There is wide discrepancy of the normal femur volume reported in two previous publications[25, 85] although identical measurement technique was used, namely multiplanar cross-sectional segmentation. One possible explanation is that this type of segmentation has poor interobserver agreement and that a different segmentation method - for instance longitudinal segmentation - may be superior. An alternative explanation is that inaccurate volume estimation may be the result of ultrasound artifact regardless of the segmentation technique. We have already discussed that artifact such as the acoustic shadow causes considerable uncertainty when trying to segment the femur (Figure 3, page 37). If this is true, then an altogether different volume estimation method may be necessary, one which overcomes the effect of acoustic shadow.

The objective of this chapter was to explore different segmentation-based techniques in order to improve measurement agreement.

3.1 Longitudinal segmentation does not improve measurement agreement

3.1.1 Introduction

Fetal femur volume measurement was first described in 2007[25]; the methodology for femoral segmentation was developed using a commercial ultrasound scan machine KretzGE Voluson 530 (Kretz, Zipf, Austria), a 3D mechanical transducer and the application GE 3D View, which is the manufacturer's software for offline analysis of volume scans. Briefly, a volume of the fetal femur is obtained using a longitudinal view as the reference plane; this is the standard 2D plane for femur length measurement. During

offline analysis, the transverse plane is selected (plane B) and the cross-sectional contour of the femur is then traced on parallel slices, from the one diaphyseal end to the other; hence this technique is known as cross-sectional segmentation. The slice density is either fixed[25] (1 mm step); or arbitrary[85], with frequent slices near the epiphyses and less dense slicing along the central diaphysis. The volume is calculated automatically by combining all traced slices. It should be noted that the above software application is not compatible with scan machines from different manufacturers.

The objectives of this study were 1) to describe an alternative segmentation technique, namely longitudinal segmentation; 2) and to compare measurement repeatability between the cross-sectional and longitudinal techniques.

3.1.2 Methods

Volume scans were obtained using a Philips HD-9 ultrasound scanner (Philips Ultrasound, Bothell, Washington, USA) with a mechanical 3D transducer (Philips V7-2). Image data files were downloaded on a PC for offline analysis. The offline image analysis and segmentation were carried out using the open-source image analysis software program MITK (Medical Imaging Interaction Toolkit MITK, version 0.12.2, German Cancer Research Center, Division of Medical and Biological Informatics, www.mitk.org).

Healthy pregnant volunteers underwent 3D scans, with informed consent and following regional ethics committee approval. One sonographer acquired one volume of the femur for each case. For cross-sectional segmentation the bone contour on plane B was traced as shown in Figure 15. For longitudinal segmentation - shown in Figure 16 - the coronal plane A was used instead; this is the reference plane of acquisition. The outline of the femur was traced on consecutive parallel slices from the one lateral border to the opposite, until the entire bone is segmented.

Each volume was segmented twice by the same operator, using the cross-sectional technique (standard method) and twice using the longitudinal technique (alternative method) in order to assess intraobserver repeatability of segmentation. A second operator

repeated the segmentations using both methods in order to assess interobserver repeatability. For each case, the repeat measurements were performed on the same volume which means that only the repeatability of the segmentation step was assessed, but not that of volume acquisition. Intraobserver and interobserver repeatability were expressed as, within-subject coefficient of variation (CV), intraclass correlation coefficient (ICC) and the Bland-Altman plots of measurement differences.

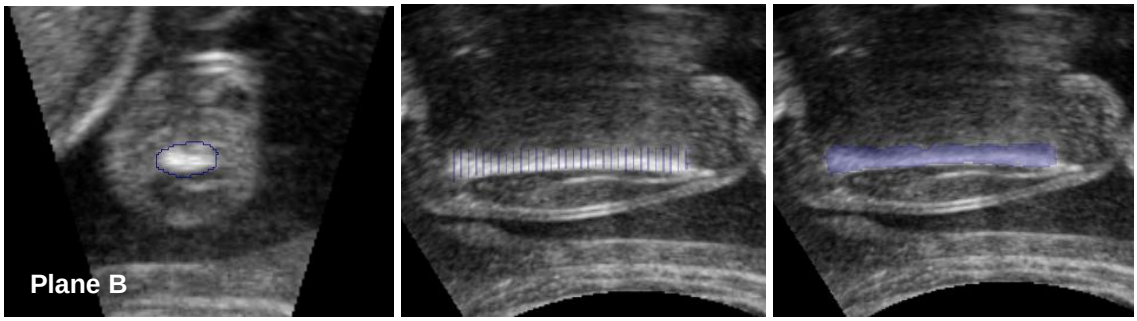


Figure 15 Technique for cross-sectional segmentation; the bone contour is traced on plane B, on several parallel slices; slices are then combined to calculate the volume

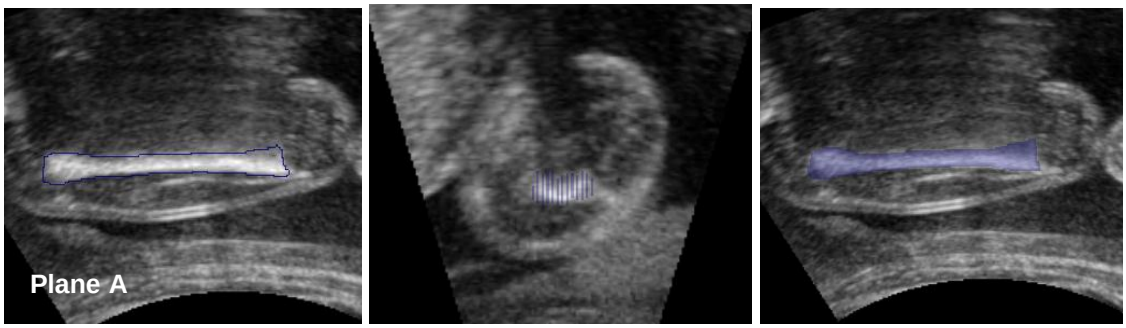


Figure 16 Technique for longitudinal segmentation; the bone contour is traced on plane A, on several parallel slices; slices are then combined to calculate the volume

3.1.3 Results

A total of 24 scans were analyzed. Median gestational age was 24 weeks (range 18 – 30 weeks). The mean number of traced slices per femur was 23.1 for the cross-sectional and

11.5 for the longitudinal technique (t-test p<0.001). CV and ICC showed that the agreement was worse when using longitudinal segmentation (Table 7). Intraobserver (Figure 17) and interobserver (Figure 18) Bland-Altman plots also demonstrated wider limits of agreement and higher systematic error (interobserver mean percentage difference = -12.1%) in the longitudinal segmentation method.

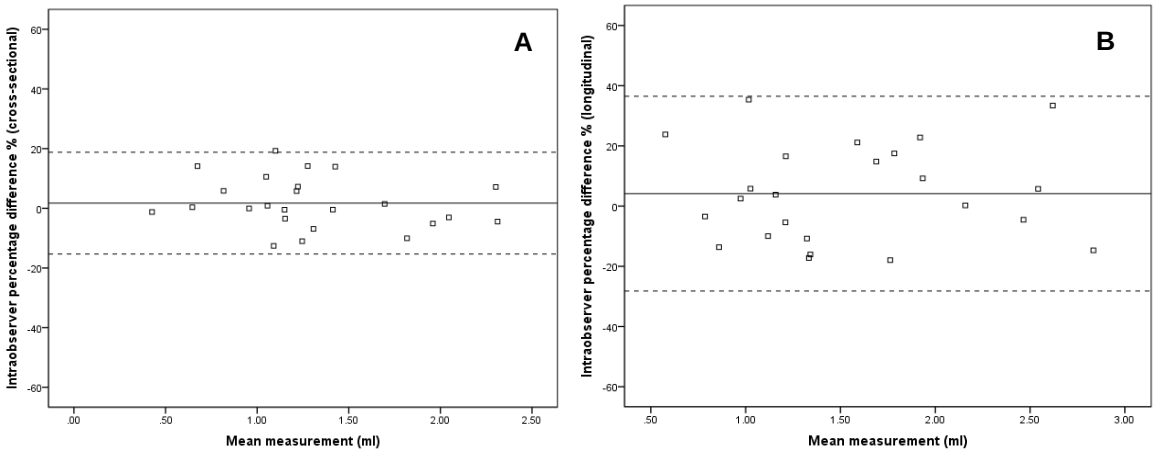


Figure 17 Bland-Altman plots of intraobserver percentage differences for cross-sectional segmentation (A) and longitudinal segmentation (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

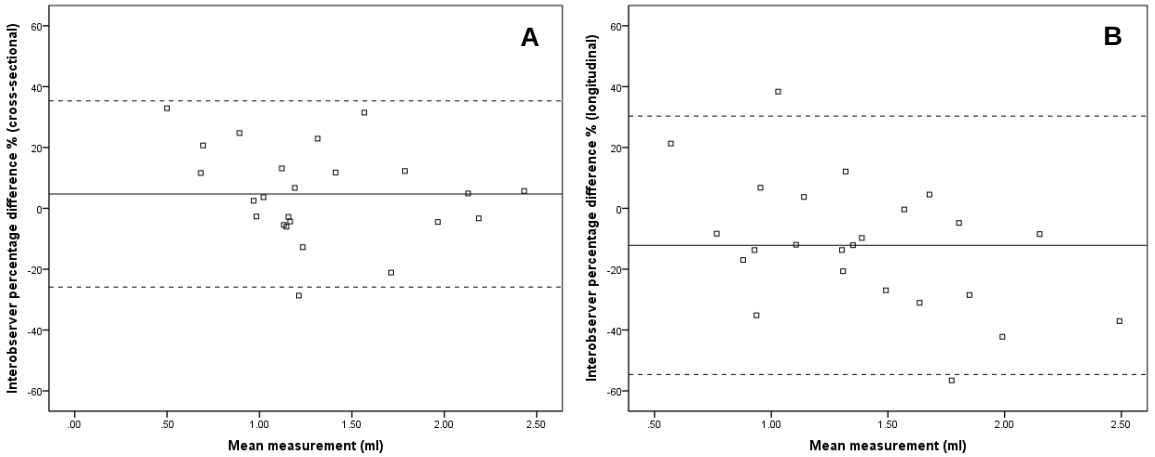


Figure 18 Bland-Altman plots of interobserver percentage differences for cross-sectional segmentation (A) and longitudinal segmentation (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

Table 7 Repeatability of cross-sectional vs. longitudinal segmentation; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval

		Cross-sectional	Longitudinal
Intraobserver	**CV**	5.8%	12.0%
	ICC (95% CI)	0.977 (0.948 – 0.990)	0.916 (0.817 – 0.962)
	Mean percentage difference	1.7%	4.1%
	95% limits of agreement	-15.3 to 18.8%	-28.2 to 36.5%
Interobserver	**CV**	10.0%	25.3%
	ICC (95% CI)	0.931 (0.849 – 0.970)	0.587 (0.189 – 0.808)
	Mean percentage difference	4.7%	-12.1%
	95% limits of agreement	-25.9 to 35.3%	-54.6 to 30.3 %

3.1.4 Discussion

This study demonstrated that - contrary to our hypothesis - a longitudinal segmentation technique has inferior repeatability compared to the cross-sectional technique. It is the first study to assess this type of segmentation.

Longitudinal segmentation is achieved using parallel coronal slices, each containing the entire length of the femur. As a result of this, a smaller number of slices need to be segmented compared to cross-sectional segmentation. This would suggest that longitudinal segmentation should be quicker, easier and more reproducible. However, although the number of traced slices is reduced, the longitudinal view of the femur suffers from considerable image artifact, which is especially evident at the outermost slices near the lateral side of the bone. The bone outline in those lateral slices is not clear and there is potential for larger error because tracing along the entire length is necessary. This was seen both in the mean measurement difference which denotes systematic error; and the limits of agreement which reflect higher random error.

This study has several strengths. For intraobserver agreement there was an interval of at least a week between first and second measurement to avoid recall bias; and for interobserver agreement a standardization exercise was initially performed and a printed

protocol was used to ensure consistency, but measurements were otherwise performed by the two observers independently.

There are also some limitations. A formal account to assess feasibility of femur volume segmentation was not undertaken, but this was not necessary as a previous report suggested that femur volume can be measured in 98% of cases when the femur can be seen on 2D ultrasound[85]. In order to ensure maximum image clarity the range of gestational ages used in this study was limited to less than 30 weeks with no cases in late gestation; however - even if later cases were included - the conclusion that longitudinal segmentation has poor repeatability would not change. Finally, there are two potential sources of measurement error, the volume acquisition and the segmentation step; the present study was designed to assess repeatability for the segmentation step only, because repeat measurements were performed on the same saved volume. Nevertheless the conclusions remain valid, because the unknown additional error associated with volume acquisition would be identical for both segmentation techniques.

3.1.5 Conclusions

Femur volume measurement using segmentation on the longitudinal plane has worse intra and interobserver repeatability than segmentation in the cross sectional plane; it is therefore not recommended.

3.2 Fusion of multiple volumes: a novel technique which can improve boundary definition

3.2.1 Introduction

Measurement error in fetal femur volumetry is possibly the result of acoustic shadow. The part of the femur which is within the “dark side” cannot be traced reliably when using

cross-sectional segmentation. In the previous experiment, an alternative longitudinal segmentation technique demonstrated poor repeatability.

One way to overcome this limitation would be to minimize the acoustic shadow at the time of volume acquisition. Unfortunately, this cannot be accomplished by altering the ultrasound machine settings such as gain, power or dynamic range. Regardless of those settings, when a single volume of the fetal femur is acquired in 3D, the voxel area situated immediately under the bone appears black. Theoretically, the amount of shadow could be minimized if it was possible to fuse several 3D volume datasets of the same femur obtained from different angles. Such image fusion techniques have been used to enhance boundary definition of the adult left heart chambers[186] and also to align and fuse multiple 4D fetal echocardiography images[187].

The objectives of this study were 1) to employ a technique called image registration and fusion in order to combine multiple 3D volumes of the same femur in one composite dataset; and 2) to assess the repeatability of femur volume measurement using this technique.

3.2.2 Methods

Volume scans were obtained using a Philips HD-9 ultrasound scanner (Philips Ultrasound, Bothell, Washington, USA) with a mechanical 3D transducer (Philips V7-2). Healthy pregnant volunteers underwent 3D scans, with informed consent and following regional ethics committee approval. For each case, one sonographer acquired six volume scans of the same femur, taken from different angles as shown in Figure 19; all six volumes included the entire femur acquired in the longitudinal view, at an angle between 0^o and 45^o to the horizontal; the central part of the ultrasound beam was over the mid shaft in two volumes; the distal metaphysis in another two volumes; and the proximal metaphysis in the remaining two volumes. All six volumes were then merged into one fused view using *image registration and wavelet-based fusion*, a semi-automated method which permits

multiple volumes of the same fetal femur to be aligned and then merged into one image dataset[188].

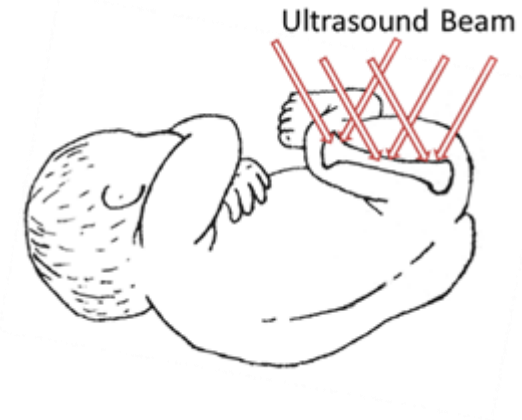


Figure 19 Six volumes of the entire femur were obtained from different angles and were fused using image registration and wavelet-based fusion

For each case, one of the single constituent volumes was randomly selected and this was compared to the fused volume. Cross-sectional segmentations of the single and fused volumes were performed twice by one operator to assess intraobserver agreement; a second operator repeated the segmentations in order to assess interobserver agreement. Intraobserver and interobserver repeatability of segmentation were expressed as, within-subject coefficient of variation (CV), intraclass correlation coefficient (ICC) and the Bland-Altman plots of measurement differences. Data were analyzed with PASW Statistics 18.0 (SPSS Inc. Chicago, Illinois 60606, USA).

3.2.3 Results

Figure 20 shows the comparison between a single dataset and the fused dataset for the same femur. On a single volume acquisition the femur appears thinner and the acoustic shadow is more obvious. When assessing the fused view it becomes evident from the outline of the ultrasound fields that several volumes were combined and stitched together in order to create such a fused dataset. The image quality is superior because the femur is better defined and “bulkier”; the acoustic shadow is reduced though still present; and

the outline of the thigh skin appears clearer all around the femur, including the area within the shadow.

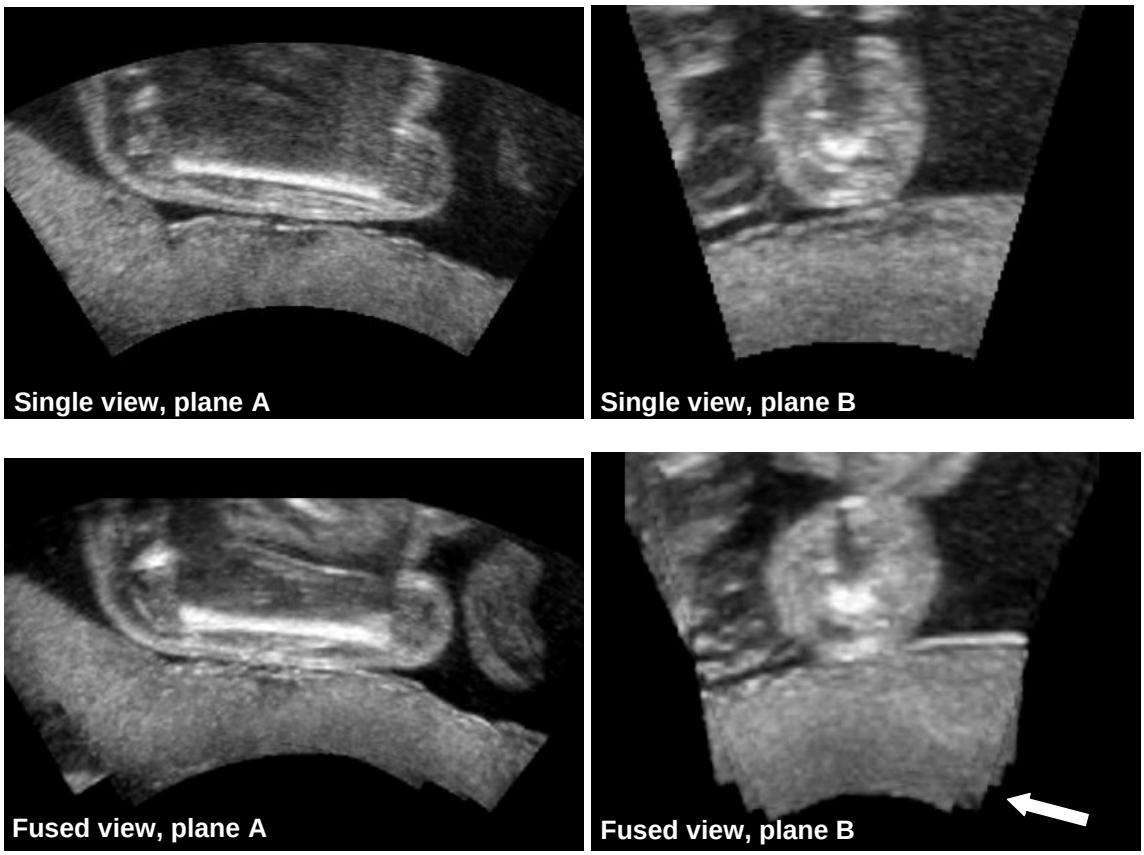


Figure 20 Images of the same femur before and after fusion; multiple ultrasound fields were combined into one fused image (arrow)

Femur fusion was attempted in 25 cases and it was successful in 24 of these. The remaining one case was rejected because of inability to align any of the six volumes. There was a median of 4 merged volumes per successful fusion, (range 2 – 6 volumes). The median gestational age was 24 weeks (range 18 – 30 weeks).

Normal volume distributions were shown on histogram inspection and using normality tests. The measured femur volume on the fused datasets was significantly increased compared to the standard single datasets: mean volume = 1.7 ml vs. 1.3 ml respectively, $p < 0.001$ (paired t-test).

Repeatability of volume measurement using cross-sectional segmentation was also improved in the fused datasets. Table 8 demonstrates the CV, ICC and limits of agreement for both intra and interobserver agreement, whereas Figure 21 and 22 demonstrate the Bland – Altman plots of percentage measurement differences.

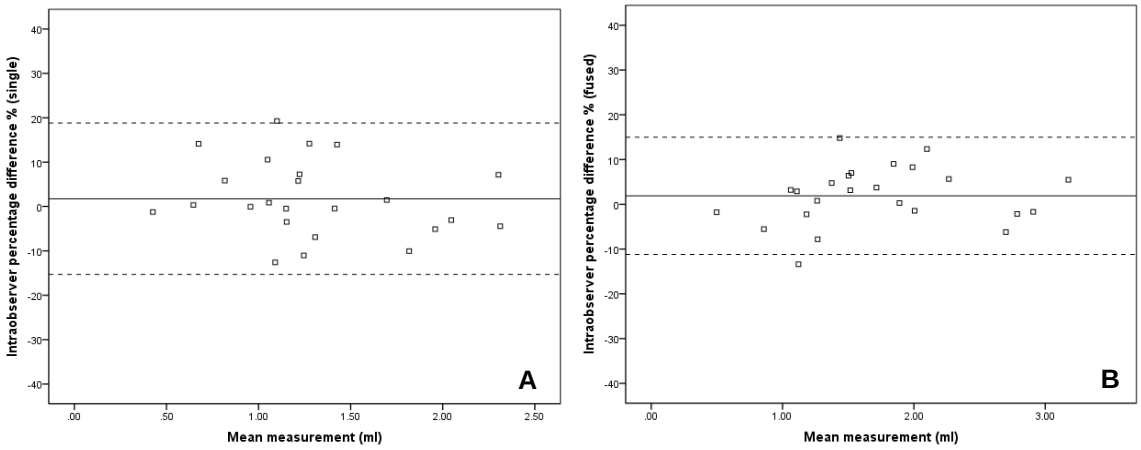


Figure 21 Bland-Altman plots of intraobserver percentage differences for segmentation on the single femoral view (A) and on the fused view (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

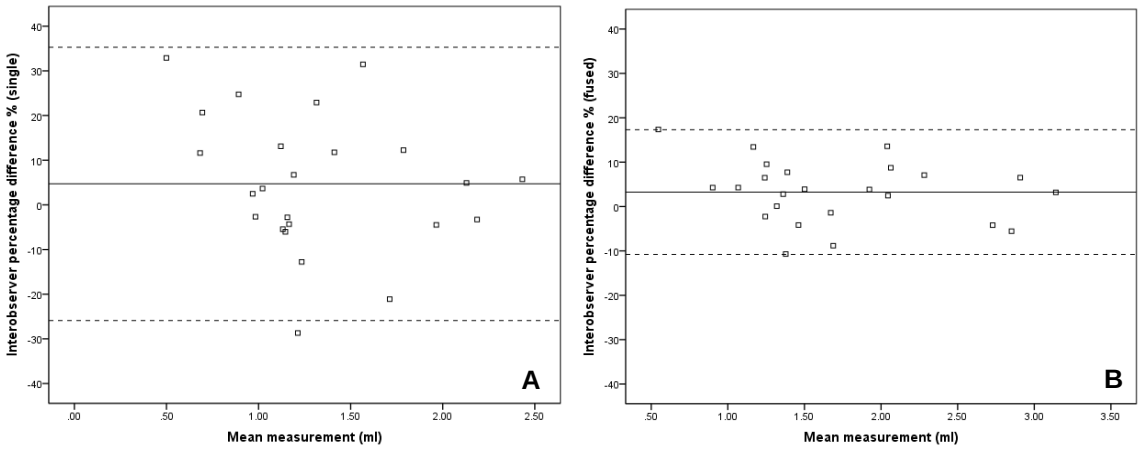


Figure 22 Bland-Altman plots of interobserver percentage differences for segmentation on the single femoral view (A) and on the fused view (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

Table 8 Repeatability of femur volume using cross-sectional segmentation on single vs. fused datasets; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval

		Single	Fused
Intraobserver	**CV**	5.8%	4.6%
	ICC (95% CI)	0.977 (0.948 – 0.990)	0.987 (0.970 – 0.994)
	Mean percentage difference	1.7%	1.9%
	95% limits of agreement	-15.3 to 18.8%	-11.2 to 15.0%
Interobserver	**CV**	10.0%	4.9%
	ICC (95% CI)	0.931 (0.849 – 0.970)	0.985 (0.964 – 0.994)
	Mean percentage difference	4.7%	3.3%
	95% limits of agreement	-25.9 to 35.3%	-10.8 to 17.3%

3.2.4 Discussion

This study has applied a novel concept of alignment and fusion of multiple 3D volumes in the assessment of fetal femur volume and has demonstrated that this technique can improve intraobserver and interobserver repeatability of cross-sectional segmentation; it also results in a modest increase of the measured femur volume.

A possible explanation for these findings is that fusion of multiple image datasets results in improved boundary definition of the fetal femur. In other words, the transition from bone to surrounding soft tissue appears clearer and therefore tracing of the bone contour becomes more reproducible. The fact that volume measurements are increased on the fused datasets further supports the hypothesis that part of the femur is usually lost in the acoustic shadow and that tracing the femur on a single 3D acquisition results in underestimation of its volume. Image fusion has the effect of casting a light on the femur from different angles which minimizes shadow and makes more of the bone “visible” to ultrasound.

There are some limitations in this study. An independent “gold standard” method for confirmation of the true femur volume does not exist here and therefore it is not clear whether this higher volume measurement is reflective of the truth. Fusion was successful

in 24 out of 25 cases which would suggest good feasibility. However this experiment was not designed to ascertain feasibility, because the case selection was not random and the range of gestational ages was narrow, with no cases in late gestation. It is possible that the feasibility of this technique in an unselected case mix may be lower. As in the previous study, repeatability of the segmentation step was only assessed, whereas the volume acquisition and fusion steps were not repeated.

Although six femur volumes were acquired from each fetus, not all of them could be successfully aligned and the final fused dataset contained a median of 4 merged volumes. In order to achieve the fusion, the image software is using the surrounding tissues as reference points in order to aid the correct alignment of the femur. It is therefore preferable that all constituent volumes are obtained within a short time interval, with the fetus lying relatively still. Gross fetal movements which alter the spatial relationship of body and limbs are likely to make alignment and fusion unsuccessful. Finally, this is a time-consuming and semi-automated technique; it relies on the extraction of multiple saved volumes onto a PC-based application and this has to be operated by a biomedical imaging engineer; registration and fusion of multiple datasets takes approximately 5 minutes and this is followed by segmentation of the fused volume which takes also approximately 5 minutes. In its present form, this technique is unlikely to be useful as a bedside real-time investigation; this however could change if the alignment and fusion procedure is developed as a fully automated facility, built into the ultrasound machine operating software.

3.2.5 Conclusions

Multiple volumes of the fetal femur can be combined in a composite 3D dataset using image registration and wavelet-based fusion. Volume measurement using cross-sectional segmentation of such a fused dataset has better repeatability; however this technique is not easily applicable as a real-time, bedside investigation.

3.3 Publications and contributions

The experiments included in this chapter were conceived and implemented jointly by the author and a DPhil student in Biomedical Image Analysis (M Yaqub). Ultrasound scans were performed by the author. Image registration and fusion was conceived and performed by M Yaqub. Femoral measurements were performed by the author and M Yaqub. Statistical analyses, figures and tables were performed by the author, except Figure 19 which was created by M Yaqub and is used with his permission. Material from this chapter was included in the following publications:

- Ioannou C, Yaqub M, Noble A, Javaid K and Papageorghiou A. Fetal femur volume measurement using fusion of multiple three-dimensional ultrasound image sets: a pilot study. Ultrasound Obstet Gynecol. 2010 Oct;36(S1): 80. doi: 10.1002/uog.8014
- Yaqub M, Ioannou C, Papageorghiou A, Javaid MK, Cooper C, Noble JA. Improving boundary definition for 3D ultrasound quantification of fetal femur. Proceedings of Medical Image Understanding and Analysis 2010: 89-94

Material from this chapter was also presented by the author under the title “Fetal femur volume measurement using fusion of multiple three-dimensional ultrasound image sets: a pilot study” at the 20th ISUOG World Congress 2010, Prague, Czech Republic (oral poster presentation); and by Dr M Yaqub under the title “Improving Boundary Definition for 3D Ultrasound Quantification of Fetal Femur” (oral presentation) at the Medical Image Understanding and Analysis 2010 Conference, Coventry, UK.

4 A new simple method for femur volume measurement

The main disadvantage of segmentation-based techniques for measurement of femur volume is that they are time-consuming and prone to systematic interobserver error. One way to make volume measurements more reproducible is to use a model which can achieve accurate volume estimation with fewer manual steps. This is conceptually similar to 2D ultrasound where, for example, manual tracing of the head circumference has been abandoned in favour of a much simpler calculation, using two intersecting diameters[189].

The objective of this chapter was to describe and validate a novel, simple method for volume calculation which obviates the need for segmentation.

4.1 In vitro validation of linear measurements on reconstructed planes

4.1.1 Introduction

It was previously discussed that tracing of the femur contour is made difficult by the acoustic shadow. In addition, the need for free-hand tracing of a large number of imaging slices makes this technique time-consuming and impractical. Obstetric ultrasonography is a real-time, bedside investigation and therefore it should be based on quick and simple measurement techniques. It is best suited to linear measurements, such as length or circumference, which form the mainstay of fetal biometry.

For femur volume calculation, linear measurements on a single 2D plane would not provide a sufficient description of the true shape of the femur. 3D ultrasound permits the reconstruction of imaging planes other than the original plane, known as plane A; for instance, the two planes which lie at 90^{o} angle to the reference plane, known as B and C planes. These reconstructed planes can be used for linear measurement in the same way as the reference plane. However, it has been anecdotally suggested that measurement on

reconstructed planes is not reliable and should be avoided. There is little evidence behind this claim. Some studies have shown that image quality when performing volumetry may be better in one or the other plane, for instance the A plane for endometrial volume[190] or the B plane for ovarian volume[191]. However these differences are not consistent and may be specific to the organ which is being assessed. It is not clear whether measurement on a reconstructed plane has the same reliability as measurement on the reference plane. The objective of this study was to compare the reliability of distance measurements between the reference and a reconstructed plane of the 3D volume, using an in vitro phantom.

4.1.2 Methods

A commercially available 3D wire ultrasound phantom (CIRS Model 055A, CIRS, Norfolk, Virginia 23513, USA, www.cirsinc.com) was used. It consists of a rigid enclosure which contains a jelly-like core of Zerdine material; this is a synthetic material with ultrasonic properties comparable to human tissue (speed of sound 1540 m/s). Within this material there are 65 parallel monofilament nylon wires of 0.1 cm diameter, at known distances from each other.

The phantom and its appearance on ultrasound are shown in Figure 23. Volumes were obtained using a Philips HD-9 ultrasound scanner (Philips Ultrasound, Bothell, Washington, USA) with a mechanical 3D transducer (Philips V7-2). Offline plane reconstruction and measurement were carried out using the open-source image analysis software program MITK (Medical Imaging Interaction Toolkit MITK, version 0.12.2, German Cancer Research Center, Division of Medical and Biological Informatics, www.mitk.org).

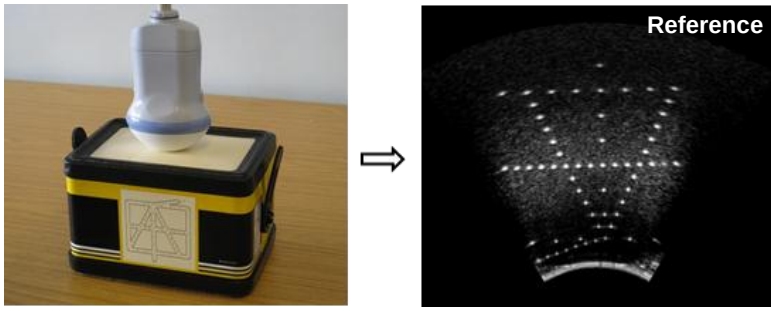


Figure 23 Wire phantom and visualization of targets on the reference plane

The technique for measurement on the *reference* plane is shown in Figure 23. The 3D transducer was positioned so that the reference plane is perpendicular the test wires; a 3D volume was acquired with a default sweep speed (5 seconds for 65^{o} sweep); when the volume was analyzed offline, the array of wire targets appeared on the reference image. A set of ten predefined distances between the targets were measured on the reference plane for each saved volume. The protocol of measurements taken is shown in Figure 24.

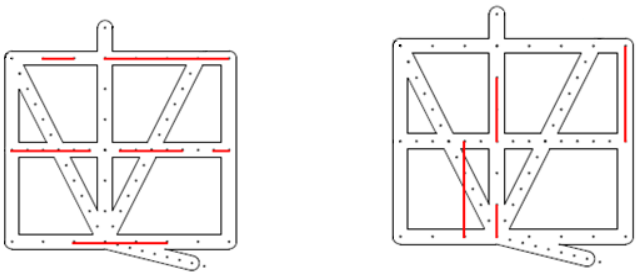


Figure 24 Measurement protocol for linear measurements on the reference image

The technique for measurement on the *reconstructed* plane is shown in Figure 25. The transducer was rotated by 90^{o} so that the reference plane becomes parallel to the test wires and a 3D volume was acquired with a default sweep speed (5 seconds for 65^{o} sweep). When the volume was analyzed offline, the array of the wire targets was not visualized on the reference plane, but on a reconstructed plane. On this reconstructed image, a set of ten predefined distances between targets were measured for each saved volume. The measurement protocol is shown in Figure 26.

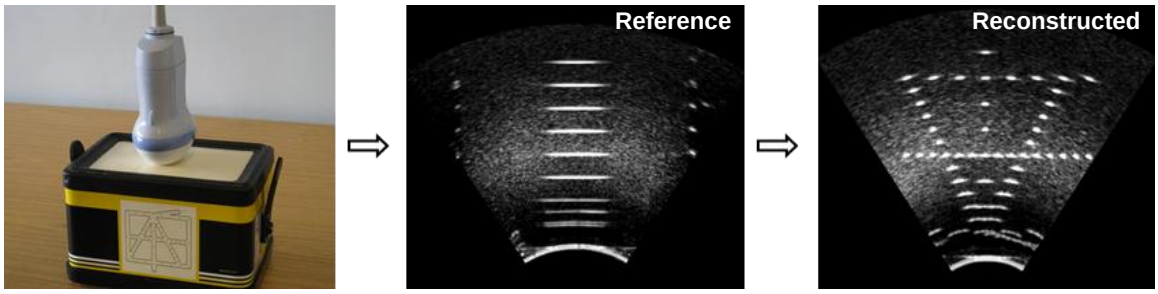


Figure 25 Wire phantom and visualization of targets on the reconstructed plane

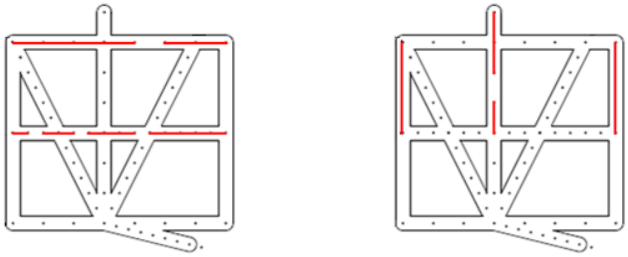


Figure 26 Measurement protocol for linear measurements on the reconstructed image

All volumes and measurements were performed by one operator. Ultrasonographic measurements were compared to the true distances between the wire targets, using Bland-Altman plots, within-subject coefficients of variation (CV) and intraclass correlation coefficients (ICC). Results were analyzed with SPSS version 19 (IBM, SPSS Statistics, US).

4.1.3 Results

There were 100 distance measurements performed on the reference plane and another 100 measurements on the reconstructed plane. The distance between targets ranged from 5 to 40 mm. Figure 27 shows a typical image of the wire targets as they appear on the reconstructed plane. This shows that there is misalignment of the targets closest to the ultrasound transducer. Only targets which are further than 3 cm from the transducer are correctly aligned on visual inspection. In order to establish whether this artifact is due to a defective transducer, a volume of the phantom and reconstruction of the wire target plane was repeated using a second Philips HD9 machine with a new V7-3 transducer.

The results were identical, with visual misalignment of the wire targets at the near field. Because of this effect it was impossible to perform any measurement within a near field of up to 3 cm, as there was no discernible target; all chosen targets for measurement were therefore situated at 3 cm or more from the transducer.

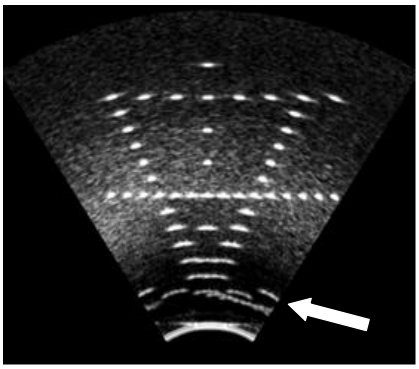


Figure 27 Reconstructed image demonstrates misalignment (arrow) of the wire targets within a short distance (3 cm) from the ultrasound transducer

The ultrasound measurements were compared to the true distance between the wire targets. For measurement on the reference plane, mean difference was -0.07 mm and 95% limits of agreement -0.57 to 0.43 mm; on the reconstructed plane mean difference was 0.11 mm and 95% limits of agreement -0.35 to 0.58 mm.

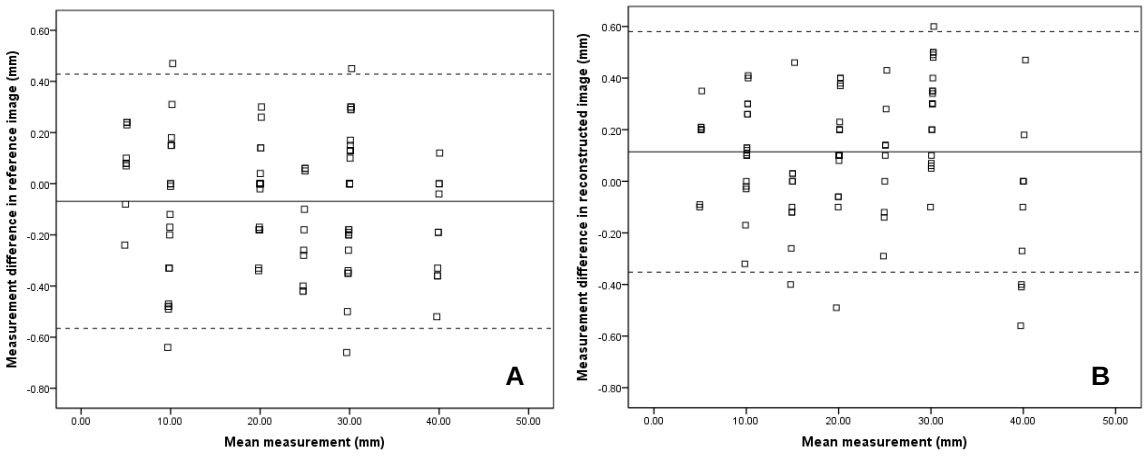


Figure 28 Bland-Altman plots of intraobserver measurement differences on the reference (A) and the reconstructed image (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

The Bland-Altman plots of measurement differences are shown in Figure 28 which demonstrates comparable limits of agreement for both imaging planes. The CV and ICC are summarized in Table 9.

Table 9 Repeatability of linear measurement on the reference and the reconstructed image; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval

	Reference image	Reconstructed image
CV	0.8%	0.9%
ICC	1.000	1.000
Mean percentage difference	-0.07 mm	0.11 mm
95% limits of agreement	-0.57 to 0.43 mm	-0.35 to 0.58 mm

4.1.4 Discussion

This study has demonstrated that plane reconstruction in 3D is associated with some distortion in the near ultrasound field, which disappears at a distance of 3 cm or more from the transducer; and that linear measurement is equally reliable in the reconstructed as it is in the reference plane.

Ten different distances were measured ten times on each volume and there was no systematic error with any of the two imaging planes (mean difference -0.07 mm and 0.11 mm respectively). The random error of measurement was approximately ±0.5 mm and this was identical in both planes. This finding demonstrates that – in ideal experimental conditions – the reconstructed plane of a 3D volume is as reliable as the reference plane and it could therefore be used for biometry.

The clinical significance of the near field distortion is uncertain. This limitation does not appear to be a manufacturing defect since it was reproduced on a different ultrasound machine of the same model specification. It is therefore an inherent limitation of this ultrasound equipment and it may not be the case with equipment from different manufacturers. Nevertheless, in the majority of cases, the transducer's near field contains

tissues of the maternal abdominal wall, such as abdominal subcutaneous fat and muscle, but no fetal parts; therefore the lack of near resolution up to 3 cm will not usually have implications in fetal biometry. It would be sensible practice for sonographers to avoid excessive pressure on the ultrasound transducer, as this may compress the abdominal tissue and may bring fetal parts nearer the “distortion zone”. This artifact should also be borne in mind whenever very slim women are scanned, where abdominal fat is minimal, especially in early gestation where the pregnancy associated edema is still absent.

4.1.5 Conclusions

Linear measurement on a reconstructed plane from a 3D volume is reliable and therefore it can be used for the purpose of fetal biometry.

4.2 A new method for femur volume measurement

4.2.1 Introduction

Although not widely used, femur volume may represent an important marker of fetal growth, both in the context of growth restriction but also as a marker of skeletal mineralization. However, volumetry using cross-sectional segmentation is subject to considerable interobserver error due to acoustic shadow. We previously demonstrated that fusion of multiple 3D acquisitions of the same femur can improve boundary definition and volume measurement reliability, but this technique has limited applicability in a busy clinical service.

One way to make volume measurements more reproducible is to use a model which can achieve accurate volume estimation with fewer manual steps, for instance a small number of linear measurements. This is conceptually similar to 2D ultrasound where, for example, manual tracing of the head circumference has been abandoned in favour of a much simpler computation, based on the assumption that the head has the approximate shape

of an ellipse so that the circumference is calculated using two intersecting diameters[189, 192]. In the case of femur volume estimation, linear measurements obtained from a single plane would not be sufficient and it is necessary to combine measurements from the reference and reconstructed planes in order to achieve accurate description of the femoral shape in 3D.

The aims of this study were 1) to propose a new mathematical model for FV calculation based on a small number of linear measurements; 2) to validate this model in vitro; 3) to establish its repeatability in vivo; and finally 4) to validate this model in vivo by comparison to postmortem computed tomography (CT) as a gold standard for FV.

4.2.2 Methods

The study consisted of four parts: describing a mathematical model to measure volume in 3D; validation using an in vitro phantom; assessment of measurement repeatability in vivo; and validation using postmortem CT scans.

Volume scans were obtained using a Philips HD-9 ultrasound machine (Philips Ultrasound, Bothell, Washington, USA) with a mechanical 3D curvilinear transducer (Philips V7-2). Static 3D volumes were loaded onto a personal computer; the offline image analysis and segmentation were carried out using the open-source image analysis software program MITK (Medical Imaging Interaction Toolkit MITK, version 0.12.2, German Cancer Research Center, Division of Medical and Biological Informatics, www.mitk.org). This application allows routine ultrasonographic measurements such as distance to be performed offline on a saved 3D volume; it also permits segmentation using the multiplanar technique.

4.2.2.1 Mathematical model

When the femur is acquired in 3D from a longitudinal view, both the reference plane A and the transverse plane B suffer from considerable acoustic shadow. On the contrary, the reconstructed C plane allows visualization of the femur without shadowing, where both the

superior and inferior boundaries are visible. The A plane is used for FL measurement according to the principles of 2D ultrasound; the C plane is then selected for measurement of the proximal (PMD) and mid-shaft diameters (MSD) as shown in Figure 29. The distinction between proximal and distal metaphysis is easy on the C plane, as the fetal buttock or knee are easily discernible. Both PMD and MSD can often be measured on the same slice; alternatively, it is sometimes necessary to scroll a few slices up or down in order to select the optimal slice for each diameter individually.

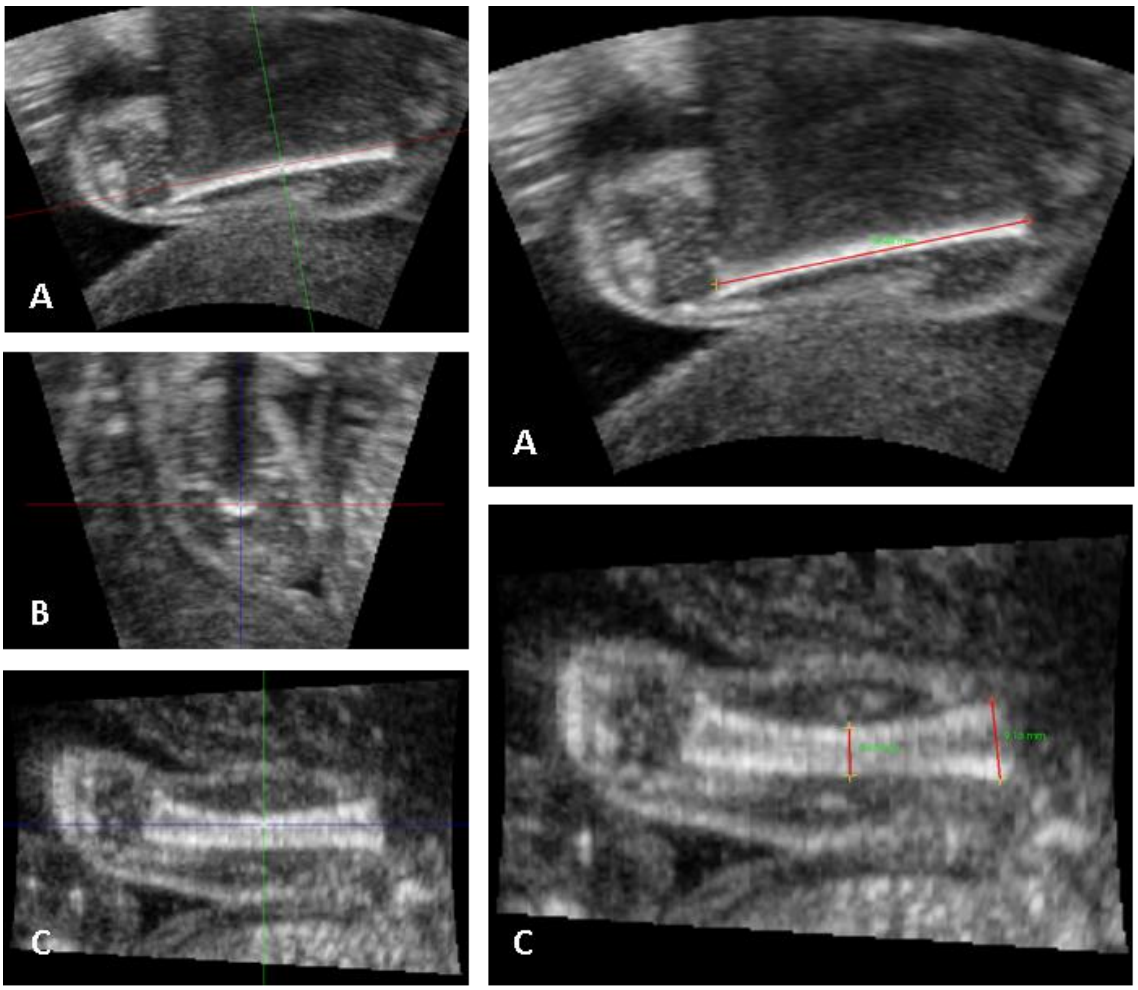


Figure 29 The femur in the three constituent imaging planes: prior to measurement the orthogonal orientation axes (shown in green, red and blue) are aligned with the longitudinal axis of the femur in planes A and C and their intersection positioned at the mid-shaft (left A, B, C). The coronal A plane is then used for FL measurement (right A) and the sagittal C plane is used for PMD and MSD measurement (right C)

These three linear measurements namely FL, PMD and MSD are used to calculate the volume. We assume that the shape of the femur is a modified cylinder which is narrower at the mid-shaft and wider at the metaphyses and that every cross-section is circular. Such a three-dimensional shape is called hyperboloid (Figure 30) and its volume FV in ml can be calculated using the PMD, MSD and FL in cm and a simple mathematical formula[193]:

$$FV = \frac{1}{3} \cdot \pi \cdot FL \cdot \left(\frac{MSD^2}{2} + \frac{PMD^2}{4} \right)$$

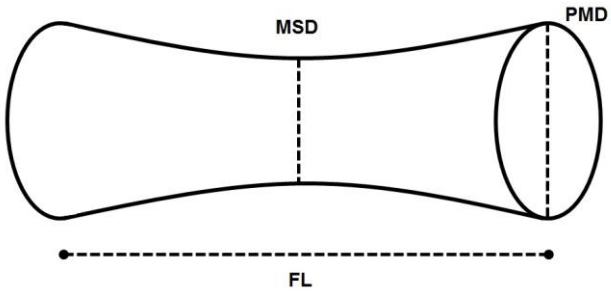


Figure 30 Schematic representation of the femur as a hyperboloid; MSD = mid-shaft diameter, PMD = proximal metaphyseal diameter and FL = femur length

4.2.2.2 In vitro phantom validation

A commercially available fetal ultrasound phantom (CIRS model 068, CIRS, Norfolk, Virginia 23513, USA) was used. This is a simulation model of a plastic fetus enclosed in jelly-like material with ultrasonic properties similar to those of a gravid uterus (Figure 31). Inside the baby's thigh, there is a femur, which appears brighter on ultrasound but does not have any acoustic shadow. Since this is an in vitro phantom, there is no fetal movement artifact. Ten volumes of the same femur were obtained by one operator who then performed volume measurements on each acquisition using both the cross-sectional segmentation technique and the mathematical model. The hypothesis was that - in this ideal experimental setting – both measurements should be identical.

4.2.2.3 Measurement repeatability in vivo

We assessed the intra- and interobserver agreement for both volumetric methods, namely the above mathematical model and the multiplanar segmentation technique[25]. Healthy pregnant volunteers with normal singleton pregnancies underwent 3D ultrasound examinations as part of a larger, observational study of fetal growth. Scans were performed with informed consent and following approval from a regional ethics committee. For intraobserver repeatability one sonographer acquired two volumes of the same femur during each scan. These were measured offline with an interval of at least one week between the first and second measurement, in order to avoid recall bias. For interobserver agreement, two sonographers scanned each woman and saved one volume each, minutes apart and independently of each other. Following this, they measured offline the saved volumes following a standardized protocol but blinded to each other's measurements. All scans and measurements were performed by three obstetricians with experience in 3D ultrasound.

Repeatability for both measurement methods was assessed using within-subject coefficients of variation (CV) and intraclass correlation coefficients (ICC). The Bland-Altman plots of intra- and interobserver measurement differences were also presented. Volume measurements and times taken per measurement for both methods were assessed with visual inspection of histogram plots and with the Shapiro - Wilk test; data were not normally distributed and were therefore compared using the Wilcoxon signed rank test. Statistical analyses and figures were generated using SPSS version 19 (IBM, SPSS Statistics, US).

4.2.2.4 Postmortem CT validation

The validity of the mathematical model was tested in a different group of women, by comparing the volume measured on ultrasound to that measured on computed tomography (CT). CT permits excellent bone visualization without acoustic artifact, and this represented the gold standard volume measurement[113]. Women undergoing planned

medical termination of pregnancy for fetal abnormality gave their informed consent to participate in this validation study, which was approved by the regional ethics committee. Inclusion criteria were a singleton pregnancy between 16 and 42 weeks of gestation, termination for fetal abnormality, maternal age above 18 years, and absence of skeletal dysplasia. A 3D ultrasound scan was performed immediately prior to the planned termination. Routine care for pregnancy termination included trans-abdominal fetocide for gestations over 20 weeks, administration of antiprogestogens and oxytocic agents to induce labour and vaginal delivery of the fetus. Following delivery, a whole-body postmortem CT was performed on a GE Lightspeed scanner (GE Healthcare, USA) using a 0.625 mm slice thickness and a fixed settings protocol. All CT scans were performed within 72 hours of the ultrasound examination. CT rendered images were created using the GE Advantage Workstation 4.4 (GE Medical Systems, USA); CT volume measurements were performed using the MITK application and the cross-sectional segmentation technique.

4.2.3 Results

4.2.3.1 In vitro phantom validation

The fetal phantom and the 3D ultrasound image of its femur are shown in Figure 31.



Figure 31 Phantom fetus and the femur in 3D ultrasound

In ten 3D acquisitions of the same femur, the volume measurements (Table 10) were identical when using segmentation (mean volume = 2.22 ml) and the novel mathematical model (mean volume 2.23 ml), mean difference = -0.01, paired t-test p = 0.884.

Table 10 Volume measurements of a phantom femur using the mathematical model and segmentation technique

Acquisition	Mathematical model (ml)	Segmentation (ml)
1	2.08	2.18
2	2.36	2.37
3	2.41	2.16
4	2.43	2.11
5	2.30	2.20
6	1.96	2.08
7	2.20	2.32
8	2.26	2.29
9	2.12	2.31
10	2.19	2.20
Mean	**2.23**	**2.22**

4.2.3.2 Measurement repeatability in vivo

FV measurements were performed on 79 fetuses, using both the mathematical model and the cross-sectional segmentation technique. Volumes derived from the mathematical model were significantly greater than those from segmentation; median FV = 2.6 ml vs. 1.6 ml respectively (Wilcoxon signed rank test p <0.001). Volume measurement using the mathematical model was significantly faster than standard segmentation: the median time to measure and record the result using the mathematical model was 1.3 min vs. 4.0 min for segmentation (Wilcoxon signed rank test p<0.001).

Intraobserver agreement for both measurement methods was tested on 43 fetuses at a median gestational age of 28 weeks (range 17 – 39); interobserver agreement was assessed on 36 different fetuses at a median of 29 weeks (range 17 – 40). For the mathematical model, mean intra- and interobserver differences were close to zero (Table 11). Conversely, the multiplanar segmentation technique showed good intraobserver

agreement but, despite initial standardization, there was high interobserver bias, with a mean percentage difference of -39.5% (Table 11).

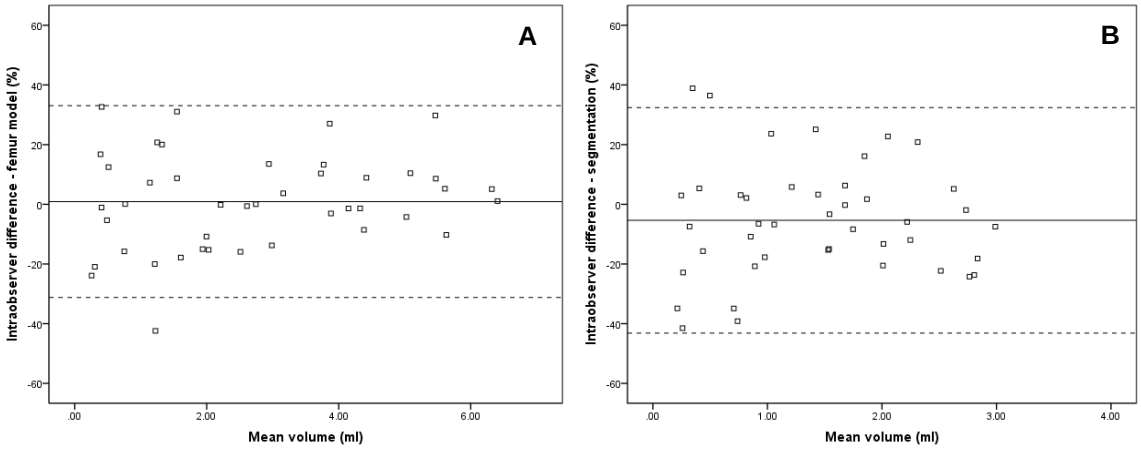


Figure 32 Bland-Altman plots of intraobserver percentage differences using the mathematical femur model (A) and the multiplanar segmentation technique (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

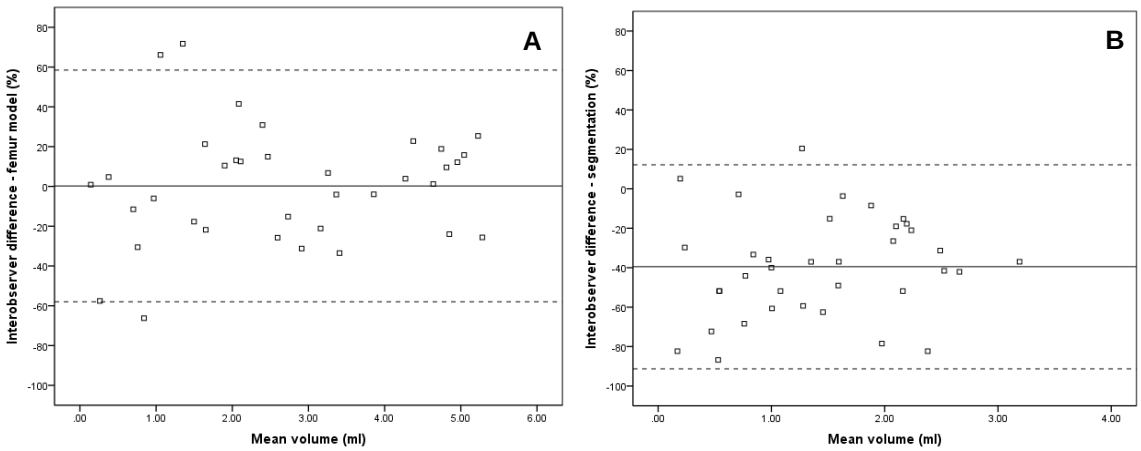


Figure 33 Bland-Altman plots of interobserver percentage differences using the mathematical femur model (A) and the multiplanar segmentation technique (B); the solid line represents the mean; dotted lines show ± 2 standard deviations

The 95% limits of agreement for the mathematical model were approximately ±32% intraobserver and ±58% interobserver; for the segmentation technique the respective

limits were ±38% and ±52% (Table 11). Figure 32 demonstrates the Bland-Altman plots of intraobserver differences for the two volumetric methods, which are similar. Conversely, the Bland – Altman plots of interobserver differences in Figure 33 demonstrate a sharp shift of the mean measurement difference into negative value, indicative of systematic error in multiplanar segmentation.

Table 11 Repeatability of femur volume measurement using the mathematical model and multiplanar segmentation; CV = within-subject coefficient of variation, ICC = intraclass correlation coefficient, CI = confidence interval

		Femur model	Segmentation
	CV	10.5%	13.2%
Intraobserver	**ICC (95% CI)**	0.977 (0.958 to 0.987)	0.952 (0.913 to 0.974)
(n=43)	**Mean percentage difference**	0.9%	-5.4%
	95% limits of agreement	-31.2 to 33.0%	-43.2% to 32.4%
	CV	16.8%	34.4%
Interobserver	**ICC (95% CI)**	0.923 (0.854 to 0.960)	0.703 (0.016 to 0.895)
(n=36)	**Mean percentage difference**	0.2%	-39.5%
	95% limits of agreement	-58.0 to 58.5%	-91.3% to 12.2%

4.2.3.3 Postmortem CT validation

A total of six cases were enrolled in the CT validation study. In all cases, skeletal dysplasias were absent and the femur length was between the 5th and 95th percentiles for gestational age. Skeletal structures were clearly defined on CT; Figure 34 shows the whole-body rendered CT of two representative cases. The femur in particular had clear demarcation, without imaging artifact. FV measurement on CT was highly reproducible, with a CV of 3.3% and ICC of 0.998. There was no statistically significant difference between the right and left FV on CT (mean difference 0.02 ml, paired t-test p = 0.531).

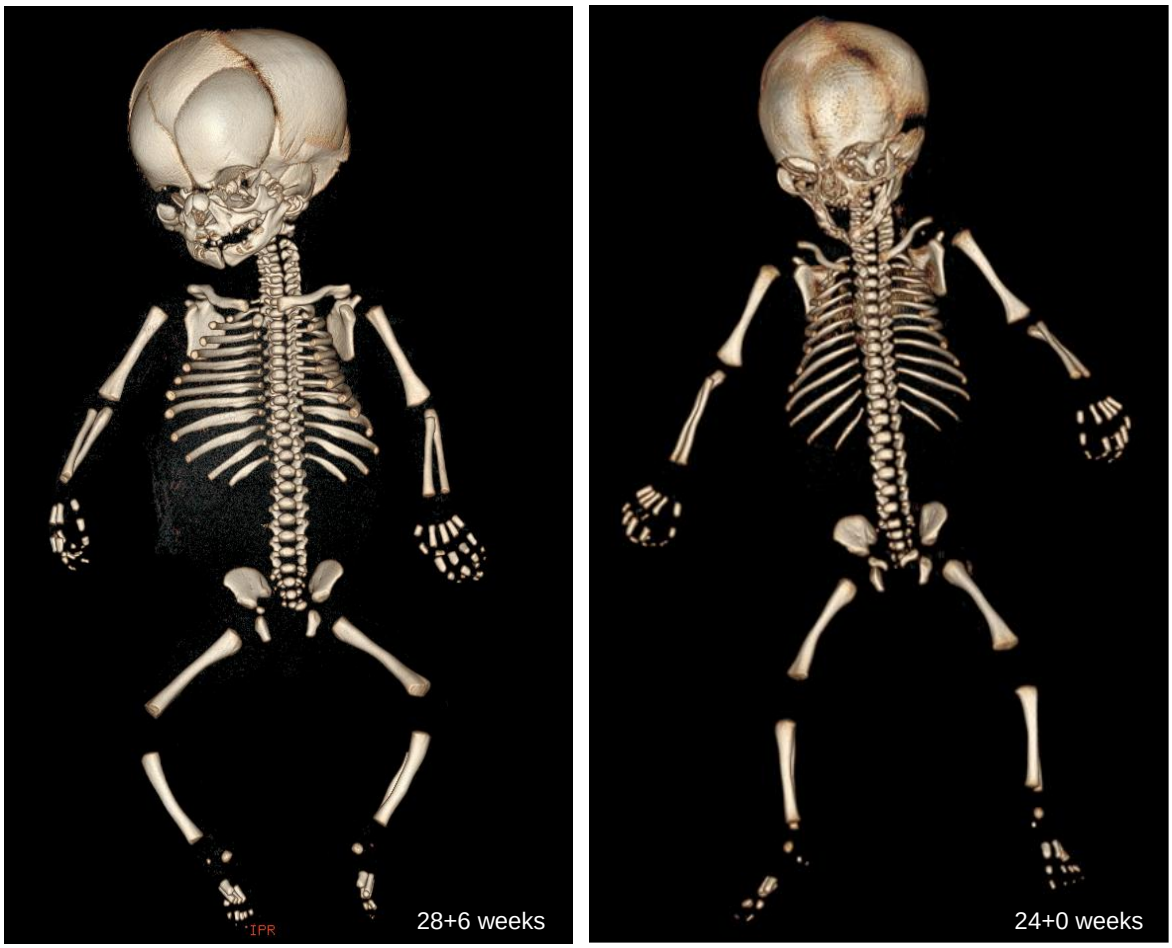


Figure 34 Two representative cases: rendered CT images at 28 and 24 weeks respectively following pregnancy termination

Femoral images were first assessed qualitatively in order to ascertain whether the femur fulfills the geometric assumptions of a hyperboloid. Visual inspection of four representative images in Figure 36 demonstrated that the longitudinal axis of the femur is not a perfectly straight line; it has a slight curvature which is more prominent along the medial femoral border and less pronounced along the lateral border. This subtle finding is also recognized in the ultrasound appearance of the femur. Although the femur length is always taken as a straight line, this slight curvature is unlikely to affect the accuracy of the mathematic model.

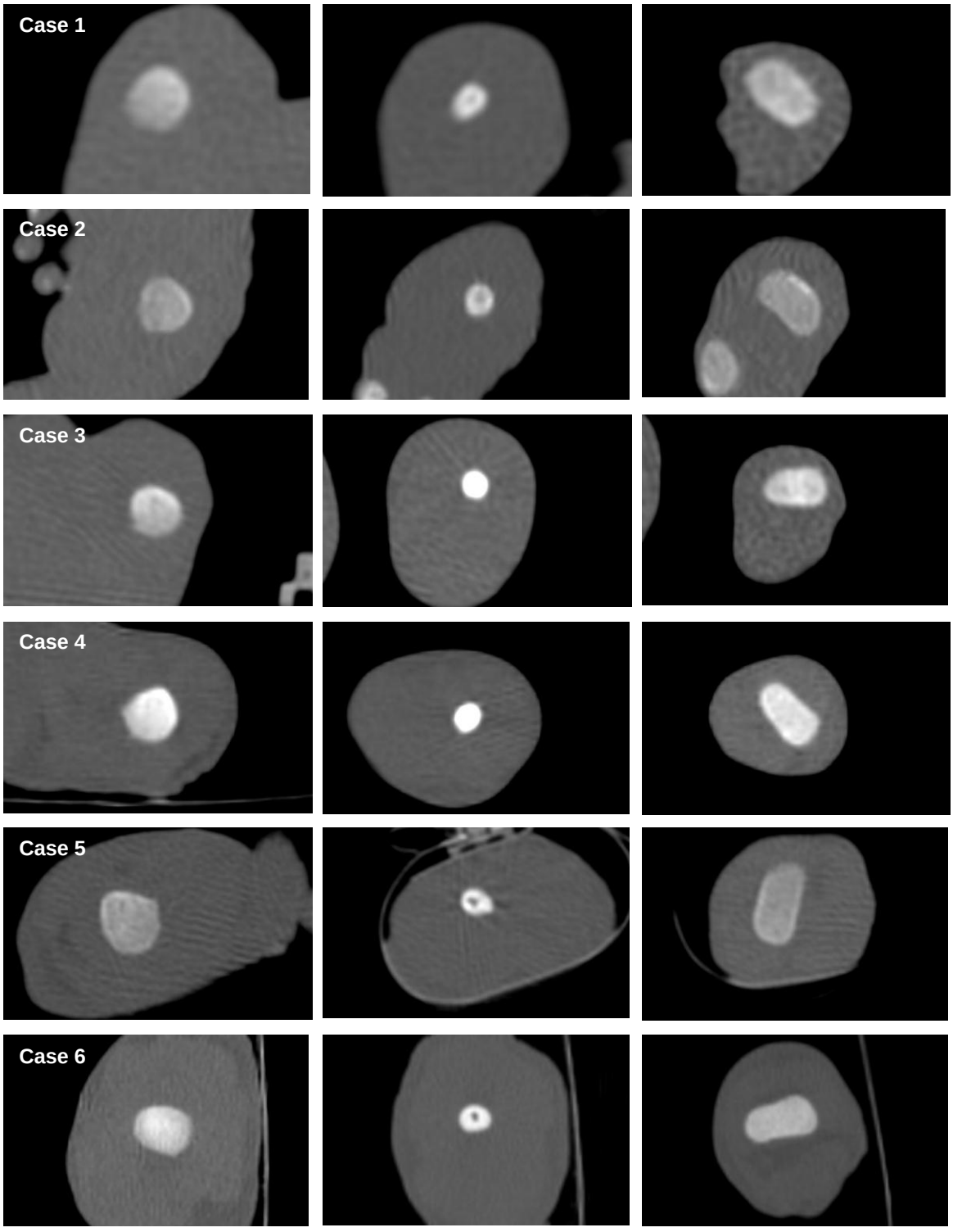


Figure 35 CT images of the fetal femur in six cases following pregnancy termination; the cross-sectional outline of the proximal metaphysis (first column) and of the mid-shaft (second column) are near circular; the outline of the distal metaphysis (third column) is flattened and ovoid



Figure 36 Longitudinal femoral curvature in four representative cases: the curvature is more prominent at the medial femoral border

Grayscale (Figure 35) and rendered CT views (Figure 37) also revealed that proximal to distal orientation is important when measuring the cross-sectional metaphyseal diameter: the two metaphyses are not identical and they should not be measured interchangeably. The *distal* metaphysis is flattened and ovoid rather than circular (Figure 37), and this would make its measurement less reproducible. It is therefore important to identify the proximal metaphysis for PMD measurement.



Figure 37 Distal metaphysis is rather flattened and ovoid: four representative cases

Finally, the ultrasound volume derived from the mathematical model was compared against the volume derived from CT. This showed excellent agreement with a mean difference = -0.01 ml which was not statistically significant (paired t-test p = 0.739). The

characteristics of included cases and the volume measurements using both techniques are shown in Table 12.

Table 12 Characteristics of cases included in the CT validation study

Case no	GA	Diagnosis	CT volume (ml)	3D US volume (ml)
1	18^{+3}	Anencephaly	0.29	0.39
2	19^{+3}	Trisomy 21	0.41	0.31
3	21^{+2}	Trisomy 18	0.37	0.40
4	22^{+6}	Trisomy 21	0.88	0.79
5	24^{+0}	Trisomy 13	1.18	1.29
6	28^{+6}	Severe brain abnormality	2.33	2.36

4.2.4 Discussion

This study described a novel, simple method for calculation of the fetal femur volume using 3D ultrasound, which is based on a set of three linear measurements and a mathematical equation. The study has shown that, compared to multiplanar segmentation, measurements using this simple technique are less prone to error; are faster to perform; and they correlate well to the gold standard of FV using CT.

The strengths of this study include the robust assessment of method repeatability and its validation using an independent gold standard. Both intra and interobserver agreement were assessed in a blinded manner and duplicate measurements were always performed on separate 3D acquisitions of the femur, rather than repeating the offline measurement of the same saved volume. Appropriate statistical methods were employed to assess measurement agreement.

Measurement error has two components: systematic error or “bias” which is expressed by the mean differences in measurements; and random error or “noise” which is demonstrated by the 95% limits of agreement. The biggest disadvantage of the multiplanar segmentation technique is the potential for large systematic interobserver bias: a systematic difference of 31% was noted between two published studies of femur volume[25, 85] despite using similar methodology. In the present study we obtained a similar

mean interobserver error of around 40% when we used the multiplanar segmentation technique. This suggests that one of the two observers systematically considered the femur to be larger than the other, and this occurred despite a standardization exercise at the outset. Review of individual volumes suggested these differences were chiefly due to the acoustic shadow: tracing on the “dark side” was variable as the femur contour is often not visible. The proposed mathematical model overcomes this limitation because the femur measurements are performed on a reconstructed plane which is devoid of acoustic shadow. As a result, the mean measurement difference between the two operators was close to zero. In addition, the above model was shown to be accurate when compared against volumes derived independently from CT.

There are some limitations in this study. There was a sizeable random error (intraobserver ±32% and interobserver ±58%), but this was similar to that seen after manual segmentation (±38% and ±52%, respectively). This imprecision was mostly due to caliper placement. “Fine” movement artifact at the time of the 3D acquisition, due to motion of the fetus over the pulsating maternal aorta, may contribute to this variability. On the sagittal plane, such fine artifact usually causes the bone outline to become less clearly demarcated and the bone edges become “fuzzy” (Figure 38). This effect may have an impact on the precision of the diameter measurements.

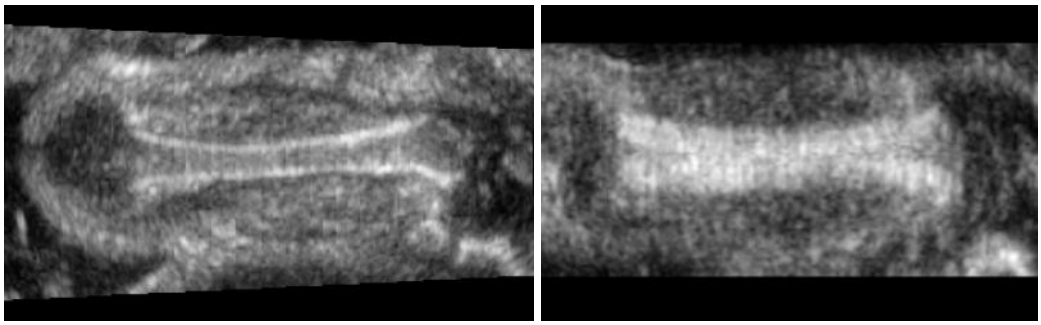


Figure 38 No movement artifact and fine artifact causing “fuzzy edges”

It is almost inevitable that measurement error of volumes - a cubic measure - would be higher than error for linear measurement. This is especially the case at earlier gestation,

when the actual volume is small, as can be seen on visual inspection of the Bland-Altman plots in Figure 32 and 33: as the volume increases, the spread of data points narrows around the mean line. Nevertheless, overall repeatability is still satisfactory in the context of biometry. In statistical terms, an ICC in excess of 0.75 denotes satisfactory reliability[194]; in the case of FV we present ICC in excess of 0.9.

The significance of measurement error should also be appreciated in the context of the actual organ size. The femur is a small structure with a volume less than 10 ml, in comparison with other fetal organs such as the fetal brain, where volume exceeds 200 ml in late gestation[51]. When the volume is larger, the same measurement noise results in smaller percentage difference and a better coefficient of variation. Conversely, in the case of the femur, the measurement noise - as a proportion of its total volume - will appear higher. In a previous study[51], brain volume calculation demonstrated intraobserver CV as high as 10%; we presented an intraobserver CV for FV calculation of 10.5% and interobserver CV of 16.8%.

4.2.5 Conclusions

We described a novel method for femur volume calculation which is suitable for clinical use. It is reliable, yet simple and quick, and can replace the more cumbersome method of manual segmentation. In addition it is associated with less systematic observer bias, even though random imprecision is similar. Finally, it is accurate when compared against a gold standard.

4.3 Publications and contributions

The author conceived and implemented all experiments included in this chapter. Ultrasound scans in healthy pregnancies were performed by the author and a second sonographer (Dr I Sarris). Offline measurements for repeatability analyses were performed by the author, Dr I Sarris and Dr R Napolitano. For the postmortem CT

validation study, ultrasound scans before pregnancy termination were performed and analyzed by the author; CT scans after termination were performed by the author and a qualified CT radiographer. The analysis of CT scans was performed by the author, assisted by a DPhil student in Biomedical Image Analysis (M Yaqub). All statistical analyses, figures and figures were performed by the author. Material from this chapter was included in the following publications:

- Ioannou C, Sarris I, Yaqub M, Napolitano R, Nicholl H, Noble JA, Javaid MK and Papageorghiou AT. A simple method for calculation of the fetal femur volume in 3D ultrasound. Ultrasound Obstet Gynecol, 40: 289. doi: 10.1002/uog.12180
- Ioannou C, Javaid MK, Mahon P, Yaqub MK, Harvey NC, Godfrey KM, Noble JA, Cooper C and Papageorghiou AT. The Effect of Maternal Vitamin D Concentration on Fetal Bone. J Clin Endocrinol Metab. 2012 Sep 18. [Epub ahead of print]

Material from this chapter was also presented by the author under the title “The effect of maternal vitamin D concentration on fetal bone” at the 22nd ISUOG World Congress 2012, Copenhagen, Denmark (short oral presentation).

5 A longitudinal study of normal femur volume

5.1 Introduction

A simple method for femur volume (FV) measurement in 3D ultrasound was described and validated in Chapter 4. Before such a method can be introduced into clinical practice, it is necessary to create a population normogram. Guidelines for studies aimed at creating fetal size charts have been proposed by various authors[195-197]. We conducted a systematic review of literature (Appendix 8.1), in order to create a comprehensive quality checklist of the appropriate study design, statistical analysis and reporting methods. Such checklist includes a prospective study design, where scans are specifically performed for the creation of a size chart; justification of sample size; accurate pregnancy dating; the inclusion of pregnancies free of conditions associated with pathological fetal growth; and appropriate statistical methods for the modeling of the raw data.

The objective of the present study was to produce a normogram for the fetal FV in a cohort of healthy pregnant women.

5.2 Methods

5.2.1 Study design

This was a prospective, longitudinal study of the normal FV using 3D ultrasound. Volume data were obtained from a cohort of women participating in a study of fetal ultrasound biometry. Women with healthy uncomplicated pregnancies were invited to participate on the basis of strict inclusion and exclusion criteria. Ultrasound scans were performed for research purposes, in order to develop prescriptive standards of fetal growth; these scans were not part of routine antenatal care or clinically indicated examinations.

From June 2009 to June 2010, consecutive eligible women were offered to take part in the study when they attended for their 12-week dating ultrasound appointment at the John

Radcliffe Hospital, UK. Following informed consent, they underwent serial ultrasound scans from 14^{+0} weeks, approximately every 5 weeks, in a longitudinal study design. This study was approved by a regional research ethics committee.

A sufficient sample size is important for precise estimation of the outer centiles of fetal size. This can be achieved using either a cross-sectional or a longitudinal study design, with appropriate analytical methods. For a longitudinal design fewer pregnant participants can be used to achieve similar precision in centile estimation, since each participant contributes more than one scan[195]. For instance, in a study of estimated fetal weight it was calculated that a longitudinal sample of 200 women can achieve similar precision in centile estimation as a cross-sectional sample of 460 participants[195]. In addition, a sample of at least 200 individuals is desirable according to the WHO Expert Committee on anthropometry[198].

Participants were dated by the first day of the last menstrual period (LMP), provided that they had a regular cycle (24-32 days), they were confident of their LMP and had a fetal crown-rump length (CRL) measurement before 14 weeks, which confirmed their dates within 7 days. Strict inclusion criteria were used in order to select pregnancies without any medical complications and without demographic or environmental factors linked to pathological growth: only naturally conceived singleton pregnancies were included, women with a BMI range of ≥18.5 and <30 kg/m^2, non-smokers, who had no pre-existing medical disease, no prior obstetric complications and no current medication or recreational drug use. In addition, the index pregnancy was prospectively monitored and clinical information was collected on research questionnaires during each scan visit and after delivery. All newborns were assessed clinically and measured within 12 hours of birth. Pregnancies were excluded from final analysis if the woman developed conditions such as gestational diabetes, gestational hypertension or pre-eclampsia; or if there was a fetal congenital or chromosomal anomaly, or intrauterine fetal death. No exclusions were made on the basis of birthweight or ultrasound measurements.

5.2.2 Measurement technique

Volume scans of the fetal femur were obtained using a Philips HD-9 ultrasound machine (Philips Ultrasound, Bothell, Washington, USA) with a mechanical 3D curvilinear transducer (Philips V7-2). A well magnified, longitudinal view of the femur in a horizontal position was used for each volume acquisition. The angle and speed of sweep were customized to fetal size and activity. The 3D volumes were downloaded onto a personal computer; the offline image analysis and segmentation were carried out using the open-source image analysis software program MITK (Medical Imaging Interaction Toolkit MITK, version 0.12.2, German Cancer Research Center, Division of Medical and Biological Informatics, www.mitk.org). All scans were performed by two obstetricians with experience in 3D ultrasound. All saved volumes were measured offline by one operator using the technique described previously.

Briefly, three linear measurements were performed as shown in Figure 29: femur length (FL) was performed on the coronal plane A; proximal metaphyseal diameter (PMD) and mid-shaft diameter (MSD) were measured on the reconstructed sagittal plane C. Measurements were done once, on a single volume per scan and with the operator blinded to the patient details and gestational age (GA). GA was calculated in days and then converted to weeks. Measurements were entered into a spreadsheet and the FV was calculated automatically using the formula[193]:

$$FV = \frac{1}{3} \cdot \pi \cdot FL \cdot \left(\frac{MSD^2}{2} + \frac{PMD^2}{4} \right)$$

5.2.3 Statistical analysis

In order to assess data for normality, the raw volume measurements were divided into six consecutive blocks of GA i.e. 14 – 18^{+6}, 19 – 23^{+6}, 24 – 28^{+6} etc. Visual inspection of histograms and Shapiro-Wilk tests demonstrated strong evidence of non-normality ($p<0.032$). Various transformations were attempted in order to normalize the data i.e. the

square root, logarithm to base 10 (log10) and the natural logarithm (ln). Following log10 and ln transformations, the data showed significant left skewing up to 24 weeks gestation ($p<0.006$). Square root transformation resulted in a normal distribution throughout gestation and was thus applied to the data.

Multilevel modeling of the median and variance was performed using MLwiN v.2.24 (MLwiN, Centre for Multilevel Modeling, University of Bristol, 2011); figures were produced using SPSS version 19 (IBM, SPSS Statistics, US). A random effects model with two levels of variance was fit to the data, where the square root of the FV was the response and GA was the explanatory variable; level 1 was defined as the intra-fetal variation on repeat scans per fetus; whereas level 2 was the inter-fetal variation, amongst different fetuses. A quadratic equation was used for modeling the median, since higher order polynomials did not improve the fit. Level 1 and 2 variances were introduced in a stepwise fashion from simpler to more elaborate models and the best fitting model was chosen by comparing deviances[199]. The total variance for the final model was calculated by adding level 1 and 2 variances[199]. The square root of the total variance was the SD and this was regressed to GA with a quadratic model providing the best fit. The fitted centile curves for the transformed data were calculated from the median using the relation: a^{th} centile = M + k SD, where M is the median (50^{th} centile), and k is the z-score corresponding to a normal equivalent deviate for the respective centile.

In order to assess goodness of fit, a standard deviation score (SDS) was calculated for every data point as follows: SDS = (actual value – fitted value) / SD; SDS scores were plotted against GA in order to identify any obvious patterns[196]. Data were back-transformed and results were presented as charts, table of median and percentile values and the regression equations for the median and percentiles. Finally, the total number of data points falling between the fitted percentile curves was calculated.

5.3 Results

Two hundred consecutive pregnant women who fulfilled the inclusion criteria were enrolled in this study. They underwent serial ultrasound scans starting from 14^{+0} weeks gestation, every 5 weeks (± 1 week) until delivery. The first ultrasound scan was arranged anytime between 14^{+0} and 18^{+6} so that follow up scans were evenly distributed across gestation. 20 women were excluded from final analysis (13 withdrew consent, 1 congenital malformation, 2 preeclampsia, 3 gestational hypertension and 1 gestational diabetes). The remaining 180 women had a total of 868 ultrasound scans which were included in the final analysis, a median of 5 scans per participant. Maternal demographic characteristics are presented in Table 13.

Table 13 Demographic characteristics of study participants; data are presented as mean (SD) or n (%)

Characteristic	Mean (SD) or n (%)
Age (years)	29.2 (4.0)
Weight (kg)	64.5 (9.4)
Height (cm)	166.5 (6.2)
BMI (kg/cm^2)	23.2 (2.8)
Parity	
0	468 (53.9%)
1	319 (36.7%)
≥2	81 (9.4%)
Gestation at delivery (weeks)	40.1 (1.3)
Birthweight (g)	3432 (461)

Multilevel modeling of the transformed data resulted in the following equations which describe the relationship between the square root of FV and gestational age (w) in weeks:

$$\text{Median } \sqrt{FV} = -1.373 + 0.127\ w - 0.000766\ w^2$$

$$\text{SD } \sqrt{FV} = 0.1234 - 0.0046965\ w + 0.0002004\ w^2$$

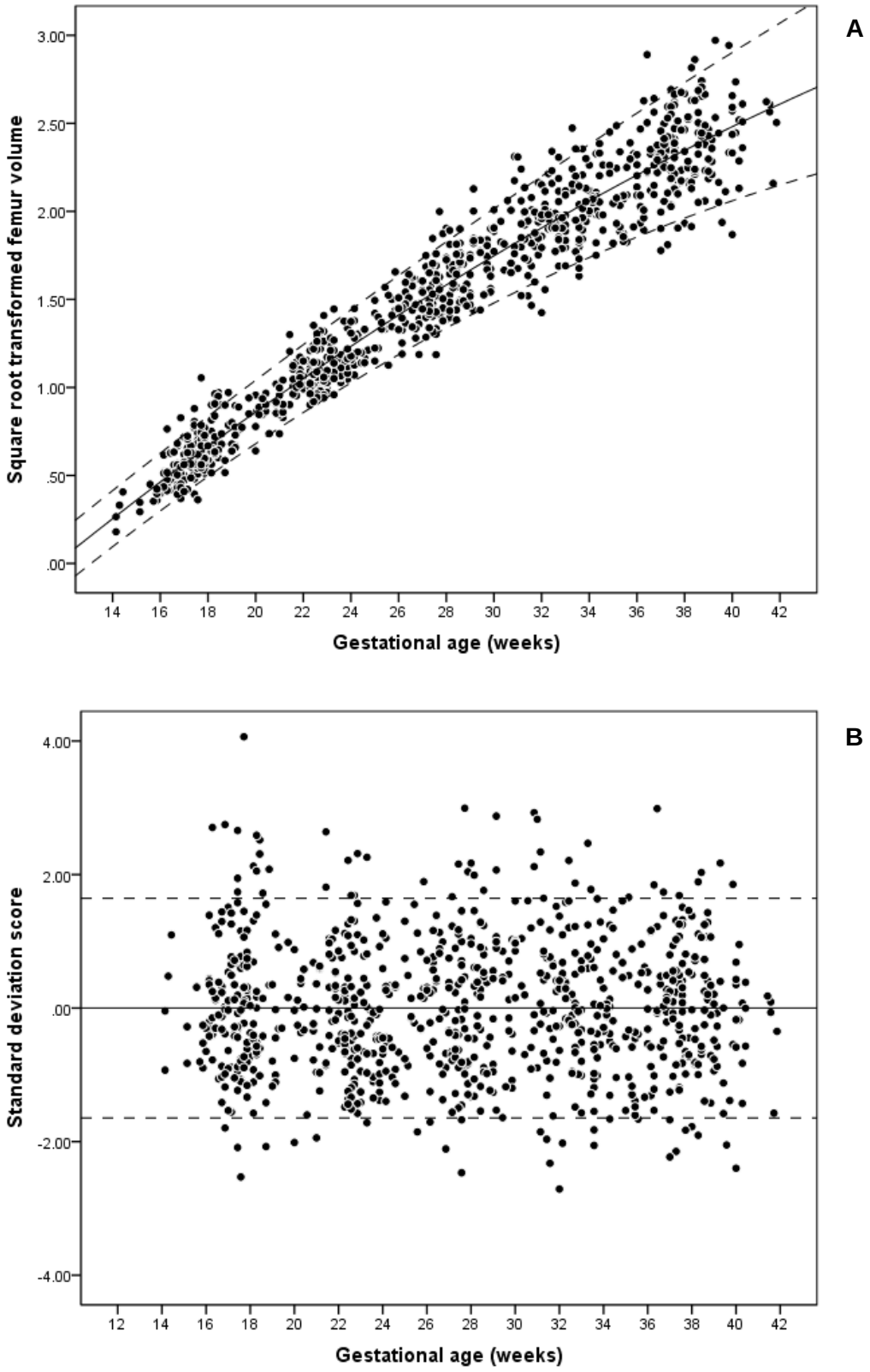


Figure 39 Scatter plot of the square root transformed data with fitted 5th, 50th and 95th percentile curves following multilevel modeling (A) and plot of SDS scores (B) demonstrating goodness of fit

A scatterplot of the transformed data is shown in Figure 39A; the fitted median, 5th and 95th percentile curves were superimposed in order to visually assess the goodness of fit. The plot of SDS scores against GA did not reveal any obvious change with gestational age (Figure 39B).

The data were subsequently back-transformed and the regression equations describing the fitted percentiles of the FV are presented in Table 14.

Table 14 Equations of femur volume (FV = ml) as a function of gestational age (w = weeks)

Fitted centile	Equation
3rd	$FV = (-1.6051269841 + 0.135834120729\ w - 0.001142868093\ w^2)^2$
5th	$FV = (-1.5760031307 + 0.134725746198\ w - 0.0010955842706\ w^2)^2$
10th	$FV = (-1.5312066952 + 0.133020915882\ w - 0.0010228553404\ w^2)^2$
50th	$FV = (-1.373 + 0.127\ w - 0.000766\ w^2)^2$
90th	$FV = (-1.2147933048 + 0.120979084118\ w - 0.0005091446596\ w^2)^2$
95th	$FV = (-1.1699968693 + 0.119274253802\ w - 0.0004364157294\ w^2)^2$
97th	$FV = (-1.1408730159 + 0.118165879271\ w - 0.000389131907\ w^2)^2$

It was calculated that 698 observations (80%) fall between the 10th and 90th centiles; 790 (91%) fall between the 5th and 95th centiles; and 818 observations (94%) fall between the 3rd and 97th centiles, confirming the goodness of fit of the model.

The median and percentile curves are presented graphically in Figure 40. The fitted values per gestational week are presented in Table 15.

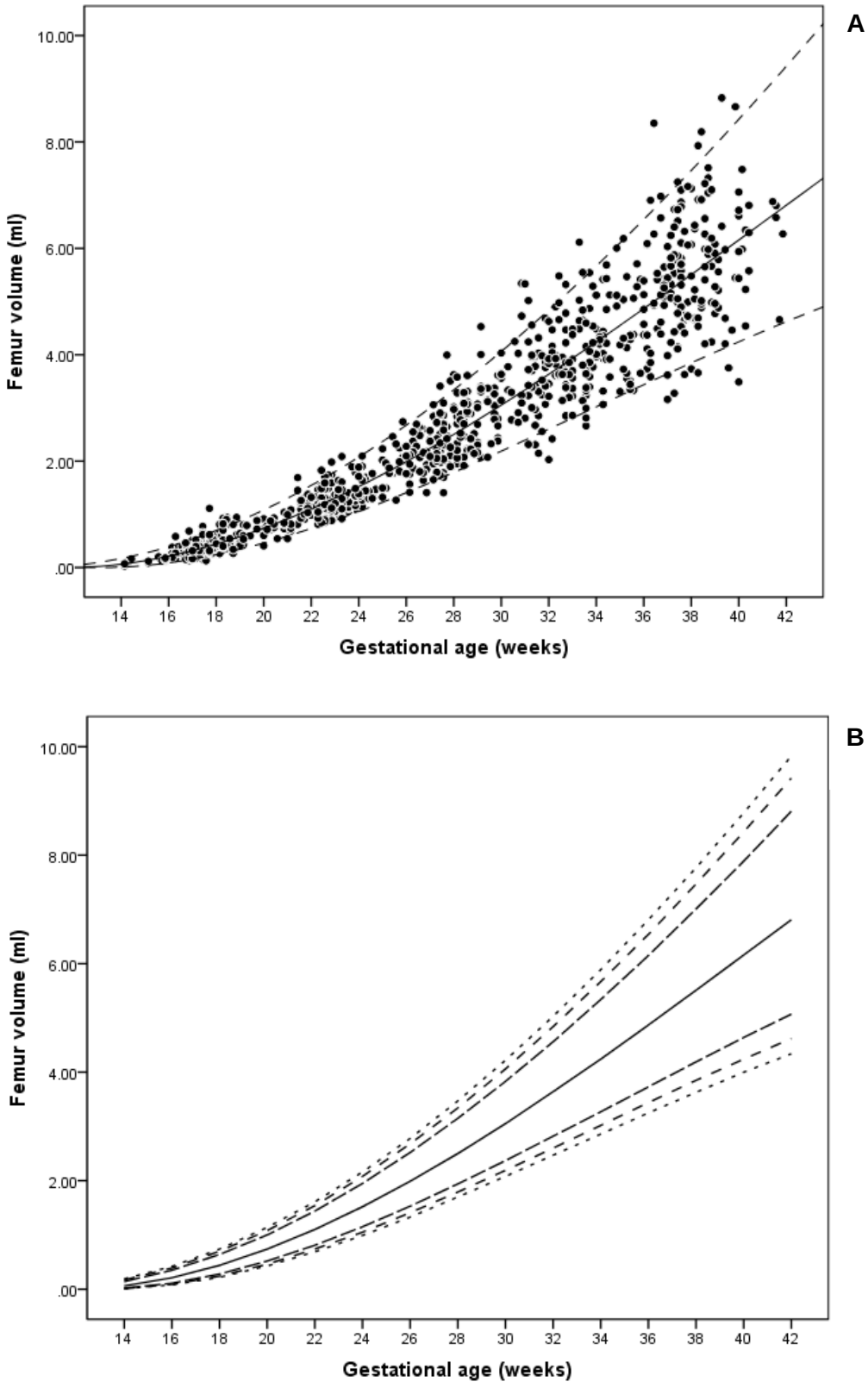


Figure 40 Fetal femur volume charts: scatter plot with superimposed 5th, 50th and 95th percentiles (A) and chart demonstrating 3rd, 5th, 10th, 50th, 90th, 95th and 97th percentile curves (B)

Table 15 Fitted median and percentile FV values (in ml) following multilevel modeling; GA = gestational age in weeks +0 days

GA	Fitted percentiles						
	3rd	5th	10th	50th	90th	95th	97th
14	0.01	0.01	0.02	0.06	0.14	0.17	0.19
15	0.03	0.04	0.05	0.13	0.24	0.27	0.30
16	0.08	0.09	0.11	0.21	0.35	0.39	0.42
17	0.14	0.16	0.19	0.32	0.48	0.54	0.57
18	0.22	0.24	0.28	0.44	0.64	0.70	0.74
19	0.32	0.35	0.39	0.58	0.81	0.88	0.93
20	0.43	0.46	0.52	0.74	1.00	1.08	1.14
21	0.55	0.59	0.66	0.91	1.21	1.30	1.37
22	0.69	0.74	0.81	1.10	1.44	1.54	1.61
23	0.84	0.89	0.97	1.31	1.69	1.80	1.88
24	0.99	1.05	1.15	1.52	1.95	2.08	2.16
25	1.16	1.23	1.33	1.75	2.22	2.37	2.47
26	1.33	1.41	1.53	1.99	2.52	2.68	2.78
27	1.51	1.59	1.73	2.24	2.82	3.00	3.12
28	1.70	1.79	1.94	2.50	3.15	3.34	3.47
29	1.88	1.99	2.15	2.77	3.48	3.69	3.84
30	2.08	2.19	2.37	3.05	3.83	4.06	4.22
31	2.27	2.40	2.59	3.34	4.19	4.44	4.62
32	2.47	2.60	2.82	3.64	4.56	4.84	5.03
33	2.67	2.81	3.04	3.94	4.94	5.25	5.45
34	2.86	3.02	3.27	4.24	5.34	5.67	5.89
35	3.06	3.23	3.50	4.55	5.74	6.10	6.34
36	3.25	3.44	3.73	4.87	6.15	6.54	6.81
37	3.45	3.64	3.96	5.19	6.58	7.00	7.28
38	3.63	3.85	4.19	5.51	7.01	7.47	7.77
39	3.82	4.05	4.41	5.83	7.45	7.94	8.27
40	4.00	4.24	4.64	6.16	7.89	8.43	8.78
41	4.17	4.44	4.85	6.48	8.35	8.92	9.30
42	4.34	4.62	5.07	6.81	8.81	9.42	9.83

5.4 Discussion

This study presents reference values for the fetal femur volume in healthy, uncomplicated pregnancies. It is the first study to employ a simple method for FV estimation based on three linear measurements and a mathematical formula.

This study has several strengths. A simple, validated 3D ultrasound technique was used. It is a prospective, observational study where ultrasound examinations were performed for research purposes and not as part of clinical care. A robust dating strategy was used in women with regular cycles, where a certain LMP was confirmed by CRL. Strict inclusion and exclusion criteria were applied in order to remove potential factors associated with pathological fetal growth. Such a strategy, often called “super normalization”, has been criticized in the past[197]. However recent opinion suggests that, when creating reference charts in anthropometry, the selection of optimal environmental factors is desirable in order to obtain “prescriptive” rather than “descriptive” standards[200].

A longitudinal study design can be used to assess fetal size, but statistical analysis should take into account the serial nature of the ultrasound data. Multilevel modeling has the advantage of being able to model the different sources and levels of variation within and between fetuses from serial ultrasound examinations[195]. Goodness of fit tests were undertaken[196] and confirmed that the resulting charts fitted the data well.

Median FV was slightly higher in the present study compared with a previous normative range published by Chang et al[25] as shown in Figure 41. Considerably smaller variance around the median was noted in the latter study, where the SD is a linear function of GA and shows only limited increase with advancing gestation. In the present study the SD is a higher order polynomial function of GA and demonstrates wider spread of observations closer to term.

There may be several reasons for such a difference. One explanation for the increased variance in the present study is higher measurement error; however our method has been validated using CT as an independent gold standard. Another explanation is that there

may be true volume differences between Taiwanese and UK fetuses. This could be possible regarding the discrepancy in median volumes. However, the differences in variance are large and would be difficult to explain biologically. The smaller variance in the study by Chang may also be the result of a smaller sample size, which has the possible effect of reducing the confidence in estimating the outer percentiles; or due to differences in the FV measurement technique.

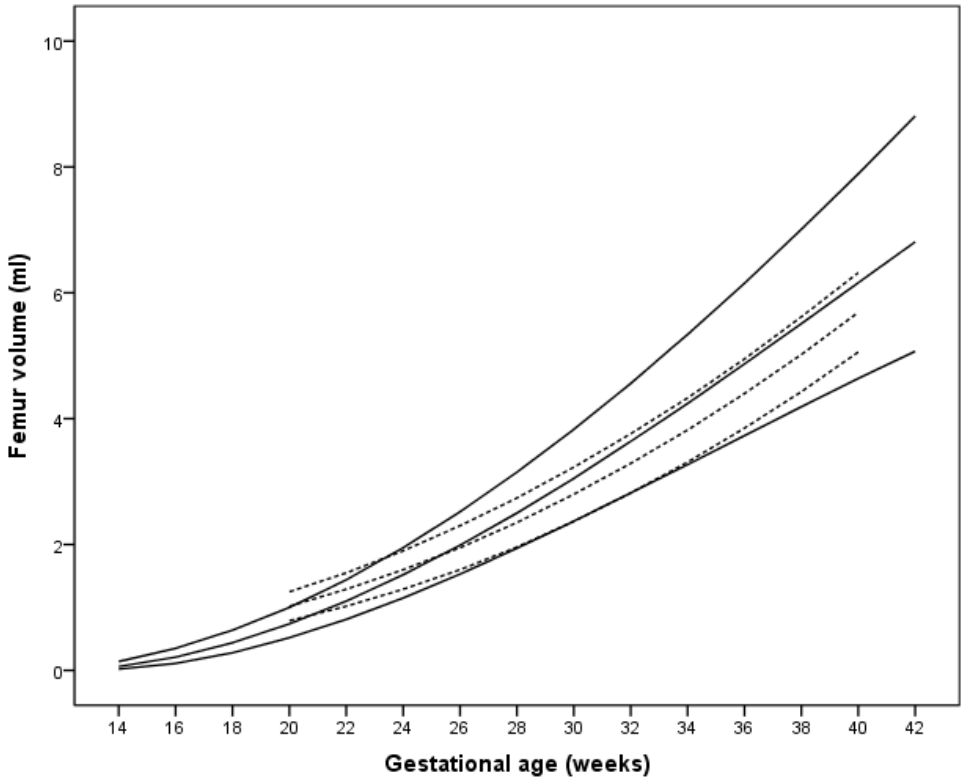


Figure 41 Fetal femur volume charts showing the 10th, 50th and 90th centiles from the present study (solid lines) compared to those from the study by Chang et al (dotted lines)

A larger increase of the SD as in our study is logical, since volume is a multidimensional measurement. A linear increase of the SD, seen with distance measures such as femur length[201], is unlikely for volume data. For instance a study of fetal lung volume[99] demonstrated a pattern of volumetric growth similar to ours: the median and SD are very

small in early gestation, whereas they both increase in a quadratic fashion towards term, giving the visual impression of a “trumpet shape” graph.

5.5 Conclusions

The present study has described a reference chart for fetal femur volume. This chart can be used as a prescriptive standard in order to assess fetal skeletal growth.

5.6 Publications and contributions

The study included in this chapter was conducted as a secondary project within INTERGROWTH 21st, an international multicenter study of normal fetal growth, which is coordinated by the University of Oxford. The overall study protocol was conceived by the INTERGROWTH 21st Steering Committee; participant recruitment, anthropometry and follow-up in the UK were implemented by a team of researchers - including the author - at the University of Oxford. Details about this study can be found at http://www.intergrowth21.org.uk/.

The femur volume project was conceived by the author. The ultrasound scans were performed by the author (485 out of 868 scans) and by a second sonographer (Dr I Sarris, 383 scans). All 868 scans were measured offline by the author. Statistical modeling, figures and figures were performed by the author with input from a medical statistician (E Ohuma).

6 The effect of maternal vitamin D concentration on fetal bone

6.1 Introduction

Maternal vitamin D deficiency during pregnancy is common in developed countries[157]; however its impact on the developing fetus is not clear. If there is a link to suboptimal skeletal growth during intrauterine life, this may be associated with a raised lifetime risk of osteoporosis, as a result of fetal programming[171]. Assessment of fetal skeletal size and growth are therefore important, yet relevant fetal biometric markers are lacking. Bone mineral content (BMC) or density (BMD) of the newborn can be assessed using dual emission x-ray absorptiometry (DEXA), but this technique is rather impractical, involves a small amount of radiation and it is only feasible after birth. Birthweight has a strong correlation with BMC[168] and is often used as a surrogate marker; however it is a crude measure which is affected not only by bone mass, but also by soft tissue growth. Prenatally, assessment of femur length (FL) using ultrasound is the most commonly used marker of skeletal biometry; it is an integral component of fetal size assessment[201] and weight estimation[202]; a short FL is also a possible marker for chromosomal anomalies[203] and fetal growth restriction[204]. However no correlation has been found between FL and maternal vitamin D levels[167].

Although not widely used, femur volume (FV) may represent an important bone marker, both in the context of fetal growth restriction but also as a specific measure of skeletal size. A reliable method for FV measurement was described and validated in Chapter 4. The objective of this study was to detect possible correlations between FV and maternal vitamin D concentration or other maternal predictors of bone size, in a cohort of healthy pregnancies.

6.2 Methods

This was an observational study of maternal vitamin D concentration, fetal FV and neonatal bone mineral indices using DEXA. We used a cohort of pregnant women from Southampton UK, taking part in the Southampton Women's Survey, an observational study of maternal nutrition and fetal growth in pregnancy, which has been described elsewhere[205]. Women aged 20 to 34 were invited to take part between 2002 and 2005; following informed consent they were interviewed by research staff at three appointments – pre-conception, early and late pregnancy - where maternal anthropometric data and lifestyle information questionnaires were collected.

A blood sample was obtained at 34 weeks for serum 25(OH)D concentration by radioimmunoassay (Diasorin, Stillwater, Minnesota, USA). This assay has a detection limit of 4 nmol/L and the analytical CV of the method is 9.1% at 22 nmol/L. Compared to the current gold-standard 25(OH)D assay i.e. the isotope-dilution liquid chromatography-tandem mass spectrometry, the Diasorin radioimmunoassay demonstrates excellent agreement in classifying individual patient samples into the correct category of vitamin D sufficiency[206].

A 3D ultrasound scan was also performed at 34 weeks by one operator, using the KretzGE Voluson 730 system (GE Healthcare, USA) with a curvilinear transducer. The 3D ultrasound femoral volumes were later measured offline by a different operator, using the method described above and blinded to all maternal and neonatal characteristics, including the vitamin D results.

Birthweight was recorded at delivery. A whole body DEXA scan of the newborn was performed within the first two weeks of age in order to assess the total BMC and total BMD, using a pencil beam densitometer Lunar DPX-L (GE Lunar, Madison, USA). A typical scan takes approximately 10 minutes and the estimated radiation dose received per examination (0.002 mSv) is only a fraction of the average daily background

radiation[207] (0.01 mSv); in comparison a typical chest x-ray generates 0.06 mSv and a CT scan up to 20 mSv[208]. All scans were performed following parental informed consent.

Univariate linear regression analyses were performed between fetal outcomes (dependent variable) and continuous maternal characteristics (independent variable). Fetal outcomes were the linear measurements of the femur (FL, PMD, MSD), FV, birthweight, BMC and BMD. Continuous maternal characteristics included maternal age, height, weight, BMI, triceps skinfold thickness (TSF) and serum vitamin D concentration. Normality of distributions was checked beforehand and logarithmic transformations (log10) were applied on the FV and TSF measurements. All ultrasound measurements were adjusted for gestational age at ultrasound; birthweight was adjusted for gestational age at delivery; and BMC, BMD were adjusted for gestational age at birth and postnatal age at DEXA. The independent samples t-test was used to assess the effect of categorical maternal variables such as parity, smoking, alcohol, exercise and season.

Following univariate analysis, all maternal predictors of the femur ultrasound measurements with a p value of less than 0.1 were fitted into a multiple regression model. Statistical analyses and figures were performed using SPSS version 19 (IBM, SPSS Statistics, US). Significance level for the multiple regression was defined as a p value of less than 0.05.

6.3 Results

The cohort consisted of 357 consecutive subjects, with a complete set of ultrasound measurements and serum vitamin D concentrations, which were performed at a mean gestation of 34.4 weeks (standard deviation SD = 0.5 weeks). Neonatal DEXA examinations were available for 203 out of 357 subjects (57%) and were performed at a mean postnatal age of 7.8 days (SD = 4.8 days).

Vitamin D deficiency (< 25 nmol/l) was encountered in 19 cases (5.3%), insufficiency (25 - 50 nmol/l) in 104 cases (29.1%) and sufficiency (≥ 50 nmol/l) in the remaining 234 cases (65.5%). Maternal and neonatal characteristics are listed in Table 16.

Table 16 Maternal and fetal characteristics; height, weight and BMI measured prior to conception; smoking and alcohol consumption documented in early pregnancy; exercise, triceps skinfold thickness, and vitamin D levels assessed in late gestation; FL, PMD, MSD, FV measured by 3D ultrasound in late gestation; BMC, BMD assessed by dual emission x-ray absorptiometry within two weeks following birth

	Characteristic	n	Median (IQR) or n (%)
Maternal characteristics	Age at delivery (years)	356	31.6 (5.3)
	Height (cm)	355	164 (8.2)
	Weight (kg)	355	63.5 (13.1)
	Body mass index (kg/m^2)	355	23.7 (4.6)
	Parity	357	
	– Nullipara		211 (59.1%)
	– Multipara		146 (40.9%)
	Smoking	353	
	– Yes		48 (13.6%)
	– No		305 (86.4%)
	Alcohol	357	
	– ≤4 units / week		267 (74.8%)
	– >4 units / week		90 (25.2%)
	Strenuous exercise	357	
	– Yes		84 (23.5%)
	– No		273 (76.5%)
	Season at ultrasound	352	
	– Winter		160 (45.5%)
	– Summer		192 (54.5%)
	Triceps skinfold thickness (mm)	347	19.9 (7.9)
	Vitamin D (nmol/l)	357	63 (42)
Fetal outcomes	Femur length (mm)	357	64.3 (3.6)
	Proximal metaphyseal diameter (mm)	357	13.6 (1.8)
	Mid-shaft diameter (mm)	357	7.00 (1.1)
	Femur volume (ml)	357	4.86 (1.21)
	Gestation at delivery (weeks)	356	40.3 (2.0)
	Birthweight (g)	356	3444 (628)
	Bone mineral content (g)	203	61.53 (18.44)
	Bone mineral density (g/cm^2)	203	0.5249 (0.033)

Maternal height and weight were significant positive predictors of almost every ultrasound parameter (Table 17). LogTSF was also a positive predictor of FL, PMD and logFV. Maternal serum vitamin D concentration demonstrated significant positive correlations with PMD and logFV. A summary of the univariate linear regression analysis between ultrasound measurements and each independent maternal variable is shown in Table 17.

Table 17 Univariate effect of maternal characteristics on fetal ultrasound measurements; p value (statistical significance) and r coefficient (Pearson correlation coefficient) derived from linear regression analysis; for categorical variables (*) the p value is derived from the independent samples t-test; FL, PMD, MSD and logFV are adjusted for gestational age at ultrasound

	FL		PMD		MSD		logFV	
	p	r	p	r	p	r	p	r
Maternal age	0.175	-0.072	0.745	-0.017	0.044	-0.108	0.193	-0.070
Height	<0.001	0.260	0.001	0.172	0.004	0.153	<0.001	0.246
Weight	0.025	0.119	0.004	0.155	0.335	0.052	0.003	0.160
BMI	0.956	0.003	0.153	0.076	0.683	-0.022	0.387	0.046
Parity*	0.756	-	0.595	-	0.156	-	0.451	-
Smoking*	<0.001	-	0.096	-	0.558	-	0.061	-
Alcohol*	0.060	-	0.345	-	0.780	-	0.760	-
Strenuous exercise*	0.266	-	0.106	-	0.388	-	0.095	-
Season at ultrasound*	0.952	-	0.232	-	0.934	-	0.405	-
LogTSF	0.018	0.128	0.010	0.139	0.884	0.008	0.013	0.134
Vitamin D	0.116	0.084	0.001	0.176	0.966	0.002	0.006	0.147

Figure 42 demonstrates the univariate effect of vitamin D concentration (divided into three groups), on femoral ultrasound measurements adjusted for GA; in Figure 43 the same data are presented in the form of linear regression scatterplots.

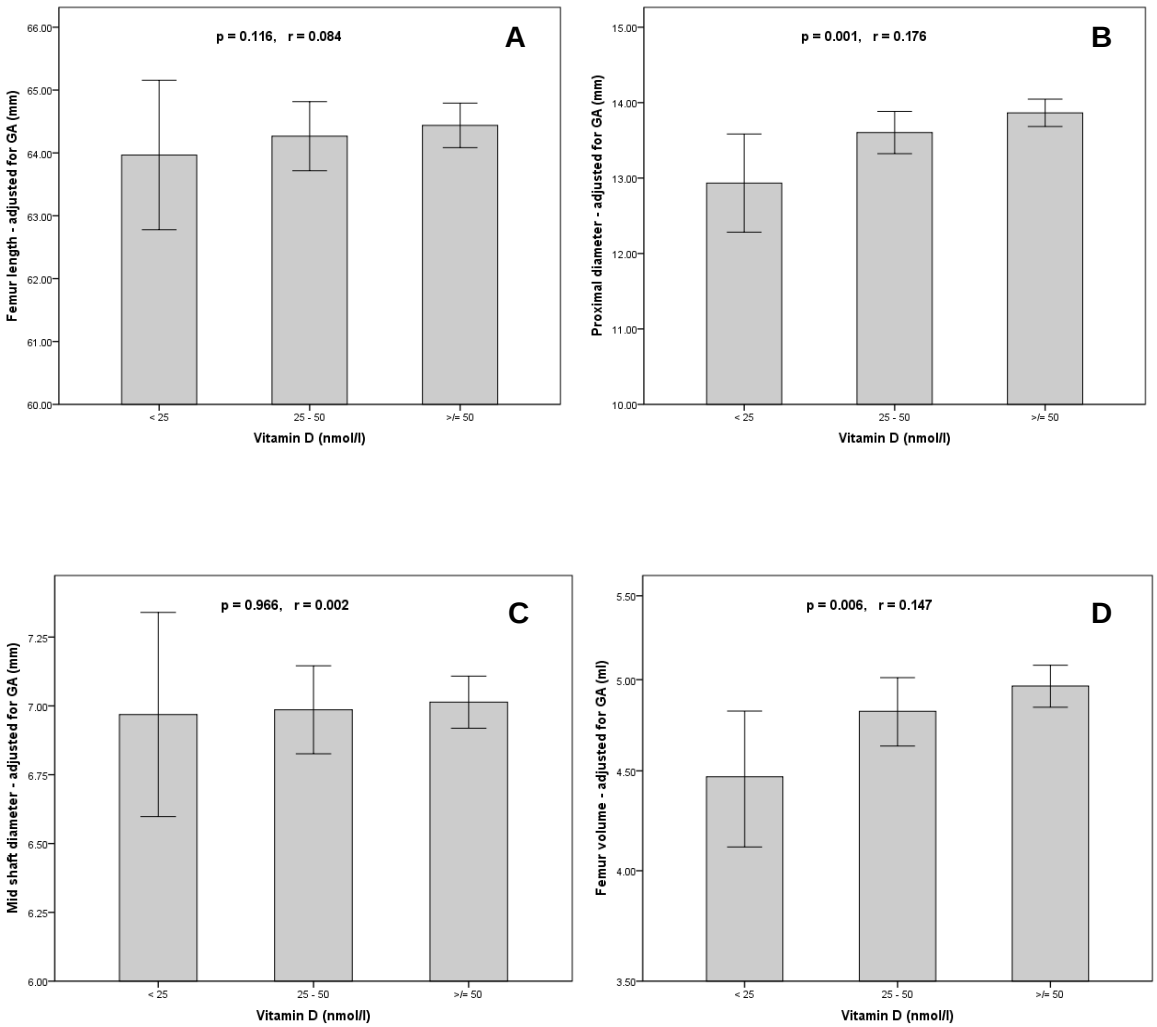


Figure 42 The effect of maternal vitamin D concentration on femur length (A), proximal metaphyseal diameter (B), mid shaft diameter (C) and femur volume (D); all ultrasound measurements are adjusted for gestational age and FV axis is shown in logarithmic scale; box plots demonstrate mean values; error bars show the 95% confidence intervals; p = statistical significance from univariate linear regression analysis; r = Pearson correlation coefficient

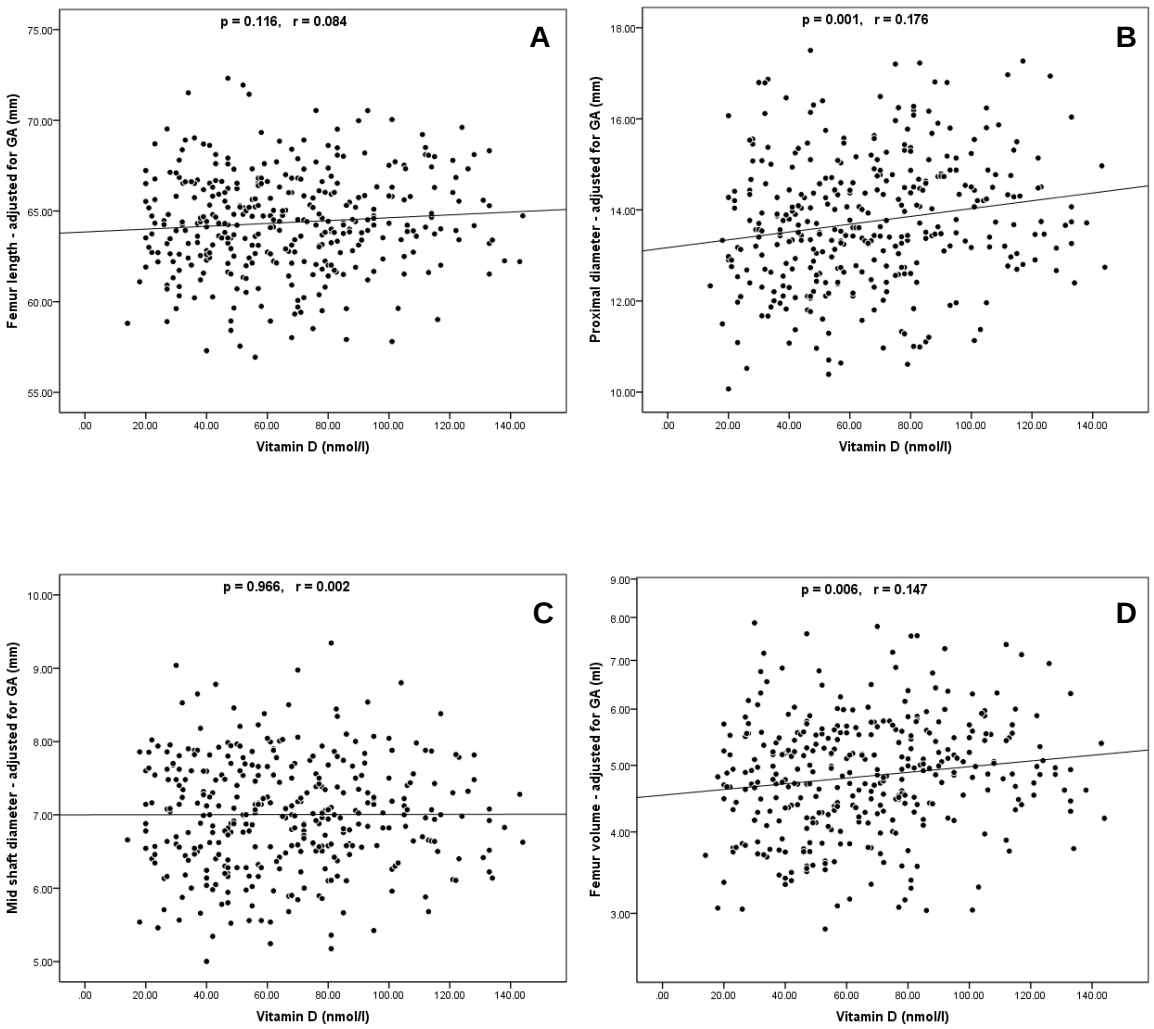


Figure 43 Linear regression scatterplots showing femur length (A), proximal metaphyseal diameter (B), mid shaft diameter (C) and femur volume (D) vs. maternal vitamin D concentration; all ultrasound measurements are adjusted for gestational age and FV axis is shown in logarithmic scale; p = statistical significance from univariate linear regression analysis; r = Pearson correlation coefficient

The univariate associations between maternal characteristics and neonatal outcomes are shown in Table 18. In summary, birthweight, BMC and BMD demonstrate no correlation with serum vitamin D concentration. BMC is correlated with maternal age, height and smoking; whereas birthweight has significant correlations with most maternal predictors of soft tissue growth (height, weight, BMI, parity, smoking, alcohol and triceps skinfold thickness).

The association between sonographic logFV and neonatal DEXA indices was assessed separately: logFV is positively correlated with BMC ($p < 0.001$, $r = 0.257$) and this persists on multiple regression analysis after controlling for the effect of maternal age, height and smoking. However, logFV is not correlated with BMD ($p = 0.461$, $r = -0.052$).

Table 18 Univariate effect of maternal characteristics on neonatal measurements; p value (statistical significance) and r coefficient (Pearson correlation coefficient) derived from linear regression analysis; for categorical variables (*) the p value is derived from the independent samples t-test; BW is adjusted for gestational age at delivery; BMC, BMD are adjusted for gestational age at delivery and postnatal age at DEXA

	BW		BMC		BMD	
	p	r	p	r	p	r
Maternal age	0.761	-0.016	0.065	-0.130	0.119	-0.110
Height	<0.001	0.248	0.096	0.117	0.453	0.053
Weight	<0.001	0.260	0.260	0.079	0.950	0.004
BMI	0.004	0.152	0.653	0.032	0.806	-0.017
Parity*	0.001	-	0.231	-	0.140	-
Smoking*	0.014	-	0.054	-	0.340	-
Alcohol*	0.048	-	0.523	-	0.048	-
Strenuous exercise*	0.708	-	0.854	-	0.619	-
Season at ultrasound*	0.611	-	0.420	-	0.987	-
LogTSF	0.008	0.143	0.107	0.115	0.563	0.041
Vitamin D	0.276	0.058	0.824	-0.016	0.095	-0.117

Multiple regression analysis using femur volume (logFV) as the dependent variable showed that the only independent predictors were maternal height and logTSF. In order to

assess which of the component femur measurements contributed to this relation, multiple regression analysis of all femur measurements was undertaken: FL demonstrated independent significant correlations with maternal height, logTSF and smoking; PMD was correlated with maternal height and vitamin D; and MSD was correlated with height and maternal age.

6.4 Discussion

This study has demonstrated that maternal height, triceps skinfold thickness and serum vitamin D concentration are significant predictors of the fetal femur size on prenatal 3D ultrasound.

Low maternal vitamin D levels are commonly encountered in developed countries: studies of pregnant women from the US[157], UK[162] and Australia[163] demonstrate a 7-18% prevalence of deficiency (<25nmol/l); whereas the combined prevalence of insufficiency and deficiency (<50nmol/l) is 33 - 49%. This is an important public health issue since it has been hypothesized that low maternal vitamin D may be associated with reduced fetal growth[163]. A number of epidemiological studies support the theory of fetal programming, according to which suboptimal bone development in pregnancy may increase the lifetime risk of osteoporosis[171]. Intrauterine life - like adolescence and puberty - is a period of intensive skeletal expansion; failure of the developing fetus to achieve its potential for bone growth could re-program its developmental trajectory.

The evidence linking reduced maternal vitamin D and fetal skeletal growth has been indirect. For instance, vitamin D deficiency has been associated with reduced birthweight[163] but not with reduced BMC of the neonate at birth[176]; however, it has been demonstrated that children born to vitamin D deficient mothers have reduced BMC measured by DEXA at the age of 9 years[162]. One explanation is that the intrauterine lack of vitamin D altered their growth trajectories, with results becoming evident several years after birth. Conversely, it is also possible that postnatal nutritional deficiencies during childhood are responsible: children share lifestyle and nutritional habits with their parents

and this may result in vitamin D deficiency during childhood, similar to maternal deficiency during pregnancy.

The present study demonstrates a direct link between maternal vitamin D and fetal bone size in utero. Univariate analysis suggests that vitamin D is a significant predictor for FV and PMD. Maternal height and TSF also have significant correlations with FL, PMD and FV. Some interesting negative correlations were also demonstrated; the negative effect of smoking on FL and FV may be explained by the overall adverse effect of smoking on fetal growth and placental function; there is also a borderline negative effect on FL from alcohol consumption; and there is a negative effect of advancing maternal age on MSD.

In order to assess the independent contributions of these factors, multiple regression analysis was performed for the ultrasound parameters. The effect of maternal weight - although strongly significant on univariate analysis - is excluded from the final model, when height and maternal adiposity (logTSF) are taken into account. The effect of vitamin D remains significant for PMD, but in the case of FV it becomes attenuated below significance. The reason is that maternal height and vitamin D are positive predictors of FV, but they are also correlated with each other ($p = 0.022$, $r = 0.121$); in other words taller mothers have higher vitamin D levels and also fetuses with larger femurs. One explanation is that the maternal height influences FV directly as a result of genetic determination and that vitamin D is merely a confounder. The alternative explanation is that in taller mothers there is increased fetal bone size as a result of the biologic effect of higher vitamin D levels.

Maternal height, adiposity and serum vitamin D concentration affect fetal skeletal size, each through a different mechanism. The first such mechanism is the effect of non-modifiable anthropometric factors like maternal height; although the association is strong its clinical value is minimal, since no intervention for optimization is possible. Secondly, bone size is indirectly regulated by the growth of the fetus as a whole; for instance increased maternal adiposity is associated with higher fetal weight and this could also lead to increased bone size. Thirdly, skeletal growth is affected by specific determinants of

bone physiology, such as maternal vitamin D, where antenatal intervention is possible. The distinction between the latter two mechanisms is not straightforward, but has an important implication: direct determinants of bone metabolism, such as maternal vitamin D, could be specifically targeted by public health interventions in order to optimize bone health. Conversely, higher maternal adiposity - although it could result in increased fetal bone size - may not always be physiological or desirable; this could only be recommended in extremely thin and underweight women.

This study has several strengths. A reliable and validated FV method was used. Several correlations between FV and maternal characteristics were demonstrated which have strong biological plausibility. Some of those correlations have been corroborated by previous studies; for instance the effects of smoking and maternal TSF have also been demonstrated in a previous study using neonatal BMC measured by DEXA[168].

Previous analysis on the same cohort demonstrated a negative correlation between maternal vitamin D and the cross-sectional area of the distal metaphysis[167]; this contradicts our finding of a positive effect of vitamin D on the proximal metaphyseal diameter. It is possible that vitamin D may have discordant biological effects between the proximal and distal metaphyseal poles. In our analysis we did not assess the distal femoral metaphysis because CT showed that this is not circular and therefore measurement of a single metaphyseal diameter could be misleading. The measurement of the cross sectional area is made difficult by the effect of acoustic shadowing, and this is also a potential source of confounding: it is plausible that higher maternal vitamin D levels result in better mineralization which in turn causes increased acoustic shadow. As a result of the shadow, the girth of the femur can be underestimated in a cross-sectional view and this may lead to the paradoxical finding that higher vitamin D is associated with a smaller cross-sectional area.

The present study has some limitations. DEXA examinations were only available in a subset of participants and this did not allow us to examine in detail the effect of maternal characteristics on neonatal indices. When interpreting the femur volume, an underlying

assumption is made that higher volume is beneficial. However size and mineralization are not necessarily synonyms; it is possible that in certain circumstances femur volume may increase in the absence of adequate mineralization and that bone “quality” may be inferior as a result. It is reassuring that, in the present data, FV has a strong positive correlation with neonatal BMC without any significant or negative effect on BMD.

Finally, the univariate effect of vitamin D on fetal femur size, although statistically significant is clinically small. Nevertheless, 3D ultrasound is a simple and safe investigation which does not carry radiation risk and is therefore likely to be acceptable to pregnant women. It can be used as a research measure in population studies or interventional trials aimed at optimizing skeletal health in pregnancy.

6.5 Conclusions

In conclusion we demonstrated that maternal height, adiposity and serum vitamin D concentration are significant predictors of femoral size. Maternal height is a non-modifiable pregnancy factor; whereas lifestyle interventions to increase maternal adiposity cannot be recommended, as the risks associated with maternal obesity would outweigh any potential benefit on fetal bone mineralization. Conversely, vitamin D supplementation is an intervention which carries minimal risk and may be beneficial for the fetus. Future trials of pregnancy supplementation should use ultrasonographic fetal femur size as an outcome measure of efficacy, which could help in establishing the optimal supplement dose; and would provide direct evidence of the potential to influence intrauterine fetal programming.

6.6 Publications and contributions

This chapter consists of an analysis of ultrasound and clinical data from the Southampton Women’s Survey, obtained through collaboration with the University of Southampton. The study protocol, participant recruitment, anthropometry, ultrasound scans and DEXA

examinations were implemented by a team of researchers in Southampton, UK. Details about this study can be found at http://www.mrc.soton.ac.uk/sws/index.asp?page=1.

The femur volume analysis was conceived by the author. All femoral ultrasound measurements were performed offline by the author, using saved 3D examinations originally performed by another sonographer (Dr P Mahon). Clinical, anthropometric and DEXA data were provided by the SWS, in database form. Statistical analyses, figures and tables were performed by the author. Material from this chapter was included in the following publication:

- Ioannou C, Javaid MK, Mahon P, Yaqub MK, Harvey NC, Godfrey KM, Noble JA, Cooper C and Papageorghiou AT. The Effect of Maternal Vitamin D Concentration on Fetal Bone. J Clin Endocrinol Metab. 2012 Sep 18. [Epub ahead of print]

Material from this chapter was also presented by the author under the title “The effect of maternal vitamin D concentration on fetal bone” at the 22nd ISUOG World Congress 2012, Copenhagen, Denmark (short oral presentation).

7 Discussion

The objectives of this thesis were to investigate the sphenoidal fontanelle area and the femur volume (FV) in 3D ultrasound as markers of fetal skeletal growth; and to correlate these with maternal vitamin D status. Although fontanelle measurement proved unreliable, a new simple and accurate method for FV was developed and validated; and a significant relationship with maternal vitamin D concentration was identified.

7.1 Summary of main findings and discussion

Observational studies of maternal vitamin D status[163] and interventional studies of vitamin D supplementation in pregnancy[138, 164, 165, 176-178] have indicated a possible link between vitamin D and fetal skeletal growth, but the evidence has been indirect and inconclusive. It would seem possible that vitamin D has either no effect or a positive effect on long bone growth (endochondral skeleton); and a possible effect of improved ossification of the cranial fontanelles (endomembranous skeleton).

One of the difficulties is how the effects of maternal vitamin D status on fetal development may be measured in order to be identified. A review of the literature on fetal biometry in 3D demonstrated that, although volumetry is reliable in vitro, there is wide discrepancy of reported normal values for the same fetal organs. This is often due to poor description of sonographic measurement methods; inadequate assessment of reproducibility, especially between different observers; and lack of validation against an independent reference method.

Therefore a large part of this research was to establish measureable manifestations on fetal skeletal health. This demonstrated that a method for fontanelle area measurement on rendered 3D ultrasound is reliable in vitro, using a custom-made ultrasound phantom. However, when this method was applied in vivo for assessment of the sphenoidal fontanelle, the measurement was not reproducible. The likely explanation is that the shape of the sphenoidal fontanelle is often irregular and lacks clear demarcation; this

makes tracing of the fontanelle prone to error. There is also another important implication: repeatability is especially poor when assessing a newly acquired image of the same fontanelle for repeat measurement. This means that images and resulting measurements taken seconds apart are often substantially different. This could either be the result of imaging artifact or true variation because of skull molding. This limitation is overlooked in the literature; studies often quote satisfactory measurement repeatability, when this is assessed with repeat measurements performed on the same "best" scan image. This has the potential to overestimate measurement agreement, since the error of repeat scan acquisition is not taken into account. In conclusion, sphenoidal fontanelle measurement was not reliable for clinical use.

The second objective of the experimental phase of this thesis was to investigate the FV as a marker of skeletal growth. An existing method for FV measurement involves tracing of the femur in parallel cross-sectional planes, but there is wide variation of results between different studies using this method[25, 85]. Two variations of this technique were explored in order to improve interobserver repeatability: a longitudinal segmentation technique was inferior in terms of intra and interobserver agreement; conversely, a method which allows the fusion of multiple 3D volumes taken from different angles, generated images with improved boundary definition and reduced acoustic shadow. FV using multiplanar segmentation in such fused image datasets, demonstrated improved repeatability and a higher volume measurement.

This experiment demonstrated that when the femur is visualized in 3D, a substantial part of the bone is lost in the acoustic shadow under its brightly reflective surface; the use of segmentation on such a standard 3D volume has poor repeatability and underestimates the true volume. Although fusion overcomes this limitation to some extent, this technique is elaborate and time-consuming; because it is not fully automated, a sonographer cannot carry out the process alone, but needs the assistance of a biomedical imaging engineer; finally, fusion is not always successful but depends on fetal activity and overall image

quality. For these reasons it was deemed necessary to develop a different measurement technique, which would combine simplicity and reliability.

A novel simple method for FV on 3D ultrasound was described and validated, one which consists of three linear measurements and a volume equation; femur length (FL) is taken on the reference plane A, whereas the proximal metaphyseal (PMD) and mid-shaft (MSD) diameters are taken on the reconstructed plane C. To achieve this, it was first demonstrated that linear measurements are equally reliable when performed either on the reference or a reconstructed plane. This mathematical model was then successfully tested in vitro on an ultrasound phantom simulating the fetal femur. Intra and interobserver repeatability of FV was assessed in vivo using both the new model and the existing segmentation method; it was demonstrated that - contrary to segmentation - the proposed model shows satisfactory interobserver agreement with no evidence of systematic error. Finally, this model had excellent agreement with the FV measured on computed tomography, in six fetuses following pregnancy termination.

The next step following validation was to apply this method on a cohort of healthy pregnant women in order to create a FV reference chart. A longitudinal study design with serial ultrasound scans was used. A normal chart was created using multilevel modeling and taking into account all methodological recommendations relevant to the creation of charts in 2D biometry. Results were presented in the form of charts, tables of median and percentile values and the respective regression equations against gestational age as the independent variable.

The final objective of this thesis was to assess FV and maternal vitamin D at 34 weeks gestation: FV had a significant correlation with maternal vitamin D, although other markers such as birthweight, neonatal bone mineral content (BMC) and bone mineral density (BMD) were not affected. It is therefore reasonable to conclude that maternal vitamin D has a positive effect on long bone size and this - given the correlation between FV and BMC - could translate to increased BMC either at birth or later in life.

7.2 Strengths

This thesis has several strengths. Fetal skeletal growth is important from a public health perspective because of the implications of fetal programming. A research hypothesis was formulated from the outset and a series of interconnected experiments were logically constructed in order to carry out the research objectives.

Evaluation of the existing literature highlighted common methodological pitfalls. In order to avoid those, the measurement methodologies used throughout this thesis were described adequately and their reliability was evaluated thoroughly; this included assessment of interobserver agreement which is often ignored in previous studies; and the error associated not only with image tracing but also the 3D volume acquisition. Most importantly, an independent gold standard was used for FV measurement, namely computed tomography.

Prior to the construction of a FV chart, a systematic review of studies in 2D biometry was carried out (Appendix 8.1) and the recommended methods were employed in sample selection, statistical analyses and data reporting. During participant recruitment, a strict selection process was used in order to minimize pathological factors which may influence growth. A number of such factors are well established; maternal smoking[209], maternal disease e.g. chronic hypertension or diabetes[210], pregnancy induced hypertension[211], pre-eclampsia[211], abnormal karyotype[210], congenital anomalies[210], pre-term delivery[212] and stillbirth[213]. Some authors have previously argued against “super normalization” i.e. applying too many exclusions, claiming that an unselected population ensures a better representation of the underlying population[197]; or that some exclusion criteria, such as maternal smoking for instance, are not reasonable[197, 214]. However, others argue that if evidence exists that any of those factors may affect fetal growth, then all should be excluded as the aim is to depict how infants *should* grow under optimal conditions (“prescriptive standard”) rather than *how they often* grow (“descriptive reference”)[200].

A number of biologically plausible correlations were demonstrated between FV and vitamin D and other maternal determinants. Such correlations were lacking in the case of other markers, such as birthweight, neonatal BMC or BMD. This is not surprising in the case of birthweight, since this is a crude measure which is confounded by the effect of soft tissue growth. Conversely, specific skeletal markers such as BMC or BMD should represent the gold standard method of assessing bone size and mineralization in the neonate. However, in the present cohort, maternal characteristics had little impact on BMC and BMD. A possible explanation is that DEXA scans in the neonatal period are technically challenging; lack of patient cooperation and excessive movement of the newborn may introduce considerable measurement error. Ultrasound FV measurement therefore represents a suitable alternative, free of radiation exposure and acceptable to pregnant women.

7.3 Limitations

This thesis has also certain limitations. The sphenoidal fontanelle was evaluated as a marker of ossification, based on the hypothesis that this fontanelle should be easier to measure on trans abdominal 3D ultrasound than the anterior fontanelle. However this hypothesis was not confirmed and although FA measurement was unreliable this does not preclude the possibility that a different fontanelle may be a useful marker; or that a different 3D visualization technique may be associated with improved measurement agreement. Measurement error is inversely related to size and therefore a larger fontanelle, such as the anterior, could be more reliable.

The proposed simple FV method reduces systematic error, yet random error is high compared with biometry in 2D ultrasound; this is an inherent weakness of measurement in multiple dimensions, as opposed to measuring distances. It is unlikely that volume measurements would be as reproducible as the biparietal diameter or the femur length. Although some reports quoting excellent repeatability in 3D can be found in the literature, we have discussed several technical pitfalls which may lead to such an overestimation.

One potential strategy to reduce random measurement error would be to obtain the average of multiple FV measurements. The downside in this approach is that it does not reflect general ultrasound practice, where the simplicity of single measurements is generally preferred. Single measurements were consistently used throughout this thesis, both in the creation of the FV chart and in the correlation analyses. Despite the “noise” associated with single measurements, the repeatability is overall still satisfactory; and numerous significant correlations with maternal characteristics are demonstrated in the last experimental chapter (Chapter 6).

This FV method is based on the inherent assumption that the shape of the femur is close to a hyperboloid. Such assumption cannot be made when studying skeletal dysplasias where the femoral shape is altered, for instance in the presence of femoral bowing. FV measurement is not recommended in the assessment of skeletal dysplasias. In addition it was assumed that the volume equation describes accurately the femur irrespective of the maternal vitamin D levels. It is possible that altered vitamin D status may induce systematic shape change so that the equation is either less accurate or biased. Although this equation was validated in six cases with postmortem CT, the maternal vitamin D status was unknown and such number of cases would be too small to deduce subtle shape variations. Nevertheless, subjective assessment of the femoral images during measurement throughout this thesis did not raise the suspicion of altered femoral shape; the individual effects of vitamin D on FL, PMD and MSD were reported and it is reassuring that no unusual pattern was identified.

There is a strong univariate correlation between vitamin D and both FV and PMD. However, following multiple regression analysis the effect on FV is attenuated, whereas the effect on PMD remains significant. The reason is that maternal height is positively associated with both vitamin D and FV. It is difficult to establish the causal pathway in this co-linear association. One explanation is that the maternal height influences FV directly as a result of genetic determination and that vitamin D is merely a confounder. The alternative explanation is that taller mothers tend to have babies with bigger bones as a

result of the biologic effect of higher vitamin D levels. Such an explanation raises the possibility that vitamin D optimization could lead to improved fetal bone growth, independently of the maternal height.

Finally it should be highlighted that that the effect of vitamin D on FV - whether real or spurious - is clinically small. The distributions of FV in vitamin D deficient vs. replete mothers have a great degree of overlap. Nevertheless, ultrasound FV is a simple investigation which does not carry radiation risk and can be used as a research measure in population studies or interventional trials aimed at optimizing skeletal health in pregnancy. As such, it should be a useful marker in trials of vitamin D supplementation.

7.4 Conclusions and implications for future research

The above findings have methodological as well as clinical implications for future research. First, there is a need for better standardization in studies of 3D biometry, with an emphasis in adequate assessment of measurement error and the need to validate novel techniques using an independent gold standard. Standardization is especially important in order to avoid systematic measurement differences in multicenter trials.

A second implication is that the proposed FV method could be useful in assessing fetuses diagnosed with excessive growth or fetal growth restriction, provided that they have a structurally normal skeleton (i.e. no skeletal dysplasia).

Finally, FV should be used as an outcome measure for studies of vitamin D supplementation during pregnancy; this may be more sensitive than existing outcomes such as birthweight or neonatal DEXA indices. If FV is increased as a result of pregnancy supplementation, this would strongly support a causal association between maternal vitamin D status and fetal bone size.

A randomized, placebo-controlled trial of vitamin D supplementation in pregnancy is needed in order to answer those questions. The Maternal Vitamin D Osteoporosis Study (MAVIDOS, http://www.mrc.soton.ac.uk/index.asp?page=284) is such a multicentre RCT, currently recruiting participants in Southampton, Oxford and Sheffield. Fetal ultrasound

parameters at 34 weeks will be used as outcome measures of efficacy. Subgroup analysis of FV would also be useful in establishing whether all pregnancies would benefit from supplementation; or whether women within a specific range of vitamin D insufficiency should be targeted. Finally, FV could also help establish the optimal supplement dose and method of administration.

In conclusion this thesis has achieved the identification of existing pitfalls in fetal skeletal imaging using 3D ultrasound; the evaluation of a measurement technique for fetal sphenoidal fontanelle measurement which was unsuccessful; the description and successful validation of a novel technique for fetal FV measurement; the creation of normal FV chart in healthy uncomplicated pregnancies; and finally the demonstration of a relationship between fetal FV and maternal vitamin D status.

8 Appendix

In Chapter 5, a normal femur volume (FV) chart from a cohort of healthy uncomplicated pregnancies was described. A new method for FV measurement in 3D ultrasound was used. Prior to the creation of this chart, we sought to identify all methodological recommendations for studies aimed at creating charts of fetal size using 2D ultrasound.

8.1 Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size

8.1.1 Introduction

Prenatal evaluation of fetal size was made possible in the 1960s by the introduction of combined A and B-mode ultrasonography in obstetrics[215]. Second and third trimester fetal biometry is now common practice and represents one of the most common medical investigations undertaken. However, a Cochrane review of trials evaluating routine ultrasonography beyond 24 weeks gestation in low risk pregnancies has not demonstrated any benefit in perinatal outcomes[216]. Furthermore, intrauterine growth restriction remains a leading cause of perinatal loss, accounting for at least one fifth of stillbirths in the UK[217]; failure to diagnose and lack of effective treatment are the likely explanations.

Recognition of pathological growth is dependent on the existence of reliable standards. However, establishing normal charts for key biometric variables such as bi-parietal diameter (BPD), head circumference (HC), abdominal circumference (AC) and femur length (FL) is not straightforward. Discrepancies of median values and percentile curves between studies have often been attributed to different distribution in racial[218], gender[219] or other biologic and demographic determinants[220]. However it is likely that differences in study design, data analysis and presentation also contribute to the observed

discrepancies. Suboptimal methodology when producing a fetal size chart is likely to affect the ability to discriminate the healthy from the compromised fetus.

Recommended methods in various aspects of study design have been published in recent years[195-197, 221, 222] including appropriate statistical methods for modelling cross-sectional[196] (one scan per fetus) or longitudinal[195] (serial scans) ultrasound data. Conversely, other aspects such as sample size, population selection and exclusion criteria remain debated to this day.

The aim of this systematic review was to evaluate the methodological quality of studies of fetal biometry by using a set of predefined quality criteria of study design, statistical analysis and reporting methods.

8.1.2 Methods

This systematic review of observational studies was conducted and reported following the checklist proposed by the MOOSE group[223]. Three major electronic databases (MEDLINE, EMBASE AND CINAHL) were systematically searched for the period 1968 to September 2011, to identify studies of two-dimensional ultrasound biometry. Reference lists of retrieved full-text articles were examined for additional, relevant citations. Studies were included if the primary objective was to create size charts for BPD, HC, AC and FL on B-mode ultrasound, in normal singleton pregnancies using either a cross-sectional or longitudinal design. The search was not restricted by study design or methodology but only articles written in English were considered. Articles were excluded if 1) only A-mode ultrasound was used, 2) the primary aim was other than the construction of size charts; for instance prediction of gestational age or comparisons between different population groups and 3) the entire gestation was not covered; for instance, size chart from 20 to 30 weeks gestation.

The keyword search strategy which was constructed by a professional information specialist is presented in Table 19. Two reviewers (KT and CI) screened the titles and abstracts of all identified citations, and selected potentially eligible studies. The full-text

versions of eligible studies were independently assessed by the same reviewers and any disagreements were resolved by consensus or consultation with a third reviewer (AP). The flow chart of the literature search is presented in Figure 44.

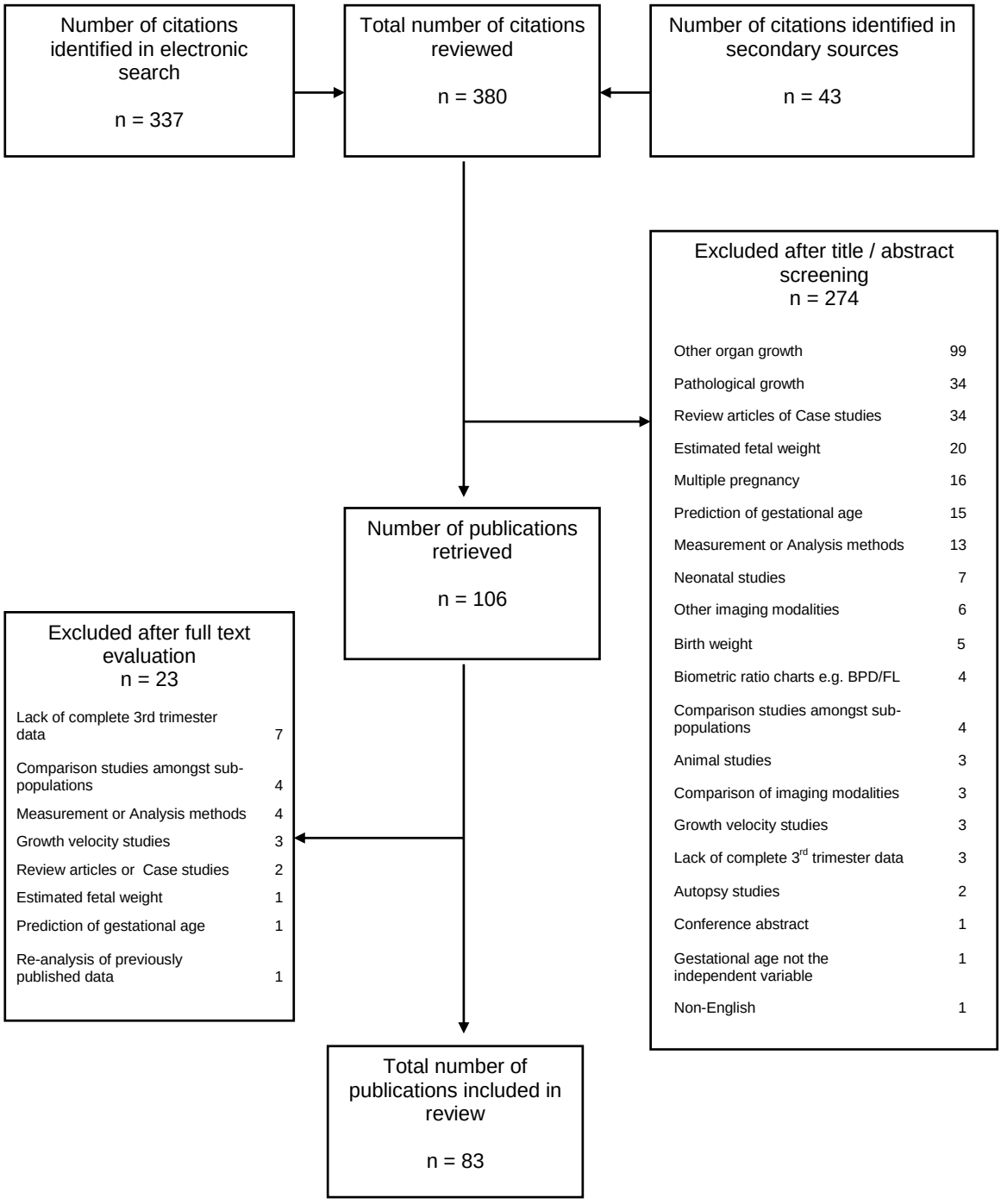


Figure 44 Consort flow diagram of literature assessment

Table 19 Search strategy

1. Fetal Development/
2. *Gestational Age/
3. ((fetal or foetal or fetus or foetus) adj2 growth).tw.
4. ((fetal or foetal or fetus or foetus) adj biometr*).tw.
5. 1 or 2 or 3 or 4
6. (growth adj (curve* or chart* or standard* or index or indices)).tw.
7. (reference adj (curve* or chart* or index or indices or equation* or value* or range* or equation* or centile* or percentile*)).tw.
8. (biometr* adj (curve* or chart* or index or indices or equation* or value* or range* or equation* or centile* or percentile*)).tw.
9. (size adj (chart* or curve*)).tw.
10. (dating adj (curve* or chart*)).tw.
11. *Reference Values/
12. 6 or 7 or 8 or 9 or 10 or 11
13. Ultrasonography, Prenatal/
14. (ultrasound* or ultrasonogra* or sonogra*).tw.
15. 13 or 14
16. 5 and 12 and 15
17. exp animal/ not human/
18. 16 not 17
19. limit 18 to english language

A list of methodological quality criteria (shown in Table 20) was initially developed by one of us (AC-A) in advance of the review and agreed between three of the authors (JV, AC-A and AP) independently of those who performed the data abstraction. These quality criteria were divided into three domains: study design, statistical methods and reporting methods; in total, 23 quality criteria were used for cross-sectional studies and 24 criteria for longitudinal studies.

Included studies were reviewed by two obstetricians (CI and KT) and a medical statistician (EO) and study details were abstracted onto an excel spread sheet. Studies were assessed against each criterion within the checklist and were scored either as “high” or “low” risk of bias. Disagreements were resolved either by consensus or consultation with a third reviewer (AP). The overall quality score was defined as the percentage of “low risk of bias” marks over the total number of quality criteria for each study.

Table 20 List of methodological quality criteria; GA = gestational age, SD = standard deviation, LMP = last menstrual period, CRL = crown rump length; for "LMS method" see Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr 1990; 44: 45-60

Domain	Low risk of bias	High risk of bias
1. STUDY DESIGN		
1.1. Design	Clearly described as either cross-sectional or longitudinal	Not reported Mixture of cross-sectional and longitudinal data
1.2 Sample selection	Population based study where there are attempts to identify and clearly define populations from a specific geographic area; from this underlying population, women are selected either consecutively or at random	Not population based; convenience sampling; arbitrary recruitment; or not reported
1.3 Number of occasions each fetus was measured (only for cross-sectional studies)	Each fetus was measured and included only once	Some fetuses were measured and included more than once
1.4. Method of selecting the gestational ages at which the fetuses were measured (only for longitudinal studies)	Interval of measures prospectively pre-specified and justified	Interval of measures not prospectively pre-specified and justified or not reported
1.5. Reason(s) for choosing a particular number of serial measurements (only for longitudinal studies)	Clear documentation of the intended number of serial measurements	No clear documentation of the intended number of serial measurements
1.6. Inclusion /Exclusion criteria	*The study made it clear that women at high risk of pregnancy complications were not included; and that women with abnormal outcome were excluded, i.e. an effort was made to include "normal" outcome as best possible* As a minimum the study population should exclude: - multiple pregnancy - fetuses with congenital structural or chromosomal anomalies - fetal death - women with disorders that may affect fetal growth (at least should specify exclusion of women with pre-existing hypertension, diabetes mellitus, renal disease and smoking) - pregnancy complications (at least preeclampsia) - pregnancies conceived by assisted reproductive technology	*The study population included both low-risk and high-risk pregnancies or women with abnormal outcome were not excluded* Study population that did not exclude fetuses or women with the characteristics previously described Exclusions which would have a direct effect on the estimated percentiles, such as fetuses found at birth to be large or small for dates
1.7. Sample size	A priori determination / calculation of sample size and justification	Lack of a priori sample size determination / calculation and justification
1.8. Data collection	Prospective study and ultrasound data collected specifically for the purpose of constructing charts of fetal size or fetal growth	Retrospective study, or data not collected specifically for the purpose of constructing charts of fetal size or fetal growth, or unclear (e.g. use of routinely collected data)

Domain	Low risk of bias	High risk of bias
1.9. Method of dating pregnancy	Clearly described Known LMP and regular menstrual cycles prior to pregnancy AND a sonogram before 14 weeks demonstrating a crown-rump length (CRL) that **corroborates LMP dates** (within how many days unspecified)	Not described clearly GA assessment at >14 weeks; or GA assessment not including USS verification
1.10. Collection of data on gestational age at inclusion	The gestational age was calculated precisely to the day	Truncation of gestational age to the number of "completed weeks"
2. STATISTICAL METHODS		
2.1. Number of measurements taken for each biometric variable	More than one measure per fetus per scan	Single measure or not specified
2.2. Statistical methods	Clearly described and identified	Not clearly described and identified
2.3. Assessment of increasing variability of the data with gestation	Performed	Not performed
2.4. Assessment of goodness of fit of the models	A test of goodness-of-fit of the models was reported	Goodness-of-fit of models was not reported
2.5. Scatter diagram of the data with the fitted percentiles superimposed	Study included scatter diagrams of the data with the percentiles superimposed	Study did not include scatter diagrams of the data with the percentiles superimposed
2.6. Change of reference percentiles across GA	Smooth change	Not smooth change
2.7. Methods used to estimate age-specific reference intervals for fetal size measurements	"Mean and SD model", smoothed crude percentiles, or "LMS method"	Inadequate (please include such inappropriate methods here)
3. REPORTING METHODS		
3.1. Characteristics of study population	Presented in a table or clearly described and includes minimum dataset of age, weight, height or BMI and parity	Not presented in a table or not clearly described, or does not contain minimum data set
3.2 Description of number approached / enrolled	Described	Not described
3.3. Ultrasound machine(s) used	Clearly specified	Not clearly specified
3.4. Number of sonographers that took the measurements	Reported	Unreported
3.5. Description of measurement techniques	The study described sufficient and unambiguous details of the measurement techniques used for fetal size parameters, including imaging plane and calliper application method	The study did not describe sufficient and unambiguous details of the measurement techniques used for fetal size parameters
3.6. Contains quality control measures	Should include the following - assessment of intra- observer variability - assessment of inter- observer variability - image review - image scoring - image storage	Does not contain quality control measures
3.7. Report of mean and SD of each measurement and the sample size for each week of gestation	Presented in a table or clearly described	Not presented in a table or not clearly described
3.8. Report of regression equations for the mean (and SD if relevant) for each measurement	Reported	Not reported

Multiple regression analysis was performed between quality scores and study characteristics which were not part of the scoring algorithm: year of publication, sample size of participating women, sample size of included ultrasound examinations, study duration, type of participating hospitals (teaching vs. non-teaching), number of participating sites (single vs. multi-site), number of sonographers (single vs. multiple) and type of country (low, middle or high income country, using the 2010 World Bank classification of economies by gross national income). Statistical analyses were performed using Microsoft Excel 2010 and SPSS version 19 (IBM, SPSS Statistics, US).

8.1.3 Results

83 publications from 32 countries were identified[11, 72, 201, 224-303]; the earliest was published in 1971 and the latest in 2008. In two publications[258, 284] multiple study designs were used, for instance AC using a longitudinal and a cross-sectional design; the best described methodology with the highest quality score was used in the final analysis. The median sample size of participating women was 558 (range 19 – 17660), whereas the median number of ultrasound examinations was 800 (range 167 – 50131). 40 studies reported one biometric parameter only, whereas the remaining 43 studies included combinations of BPD, HC, AC or FL but only 22 studies reported a complete set of all four parameters. In total there were 60 studies for BPD, 34 for HC, 41 for AC and 43 for FL.

The study characteristics and overall quality score for each study are presented in Table 21. 61 studies had a cross-sectional design and 22 studies had a longitudinal design. Amongst the 61 cross-sectional studies, only 36 (59%) clearly specified that one examination per fetus was performed during the study period. Amongst the 22 longitudinal studies, only in 12 (55%) were ultrasound data analysed using a method that took into account their serial nature.

Table 21 Characteristics of included studies (CM: cross-sectional data, several scans per fetus, C1: cross-sectional data, one scan per fetus, CS: cross-sectional data, number of scans per fetus not sated, LC: longitudinal data with cross-sectional analysis, LL: longitudinal data with longitudinal analysis)

Author	Year	Country	Number of women	Number of scans	Weeks	Measurement	Design	Data collection	Scans for research purposes only	Quality score (%)
Aickin et al[224]	1976	New Zealand	446	1018	19 - 41	BPD	CM	unclear	unclear	17
Al-Meshari et al[225]	1987	Saudi Arabia	1570	1570	12 - 41	BPD	C1	unclear	unclear	30
Amoa et al[226]	1993	Papua New Guinea	425	1229 - 1514	12 - 42	BPD, AC, FL	LC	unclear	unclear	21
Ashrafunnessa et al[227]	2003	Bangladesh	710	710	12 - 42	BPD, AC, FL	C1	prospective	yes	70
Ayangade et al[228]	1986	Nigeria	558	712	12 - 40	BPD	CM	unclear	unclear	26
Beigi et al[229]	2000	Iran	1324	15594 - 15693	12 - 40	BPD, FL	LC	unclear	unclear	13
Bergsjo et al[230]	1976	Norway	131	421	24 - 44	BPD	CM	unclear	no	13
Brons et al[231]	1990	Netherlands	63	520	12 - 40	BPD, HC, AC, FL	LL	unclear	unclear	42
Browne et al[232]	1992	USA	8285	7249 - 8108	10 - 44	BPD, HC, AC, FL	C1	retrospective	no	35
Campbell et al[233]	1971	UK	574	1029	13 - 40	BPD	CM	unclear	unclear	17
Chan et al[234]	1991	Singapore	1442	1442	17 - 40	BPD	C1	retrospective	no	22
Chang et al[235]	1996	Taiwan	2077	2077	16 - 41	BPD	C1	unclear	unclear	52
Chitty et al[236]	1994	UK	610	425 - 610	12 - 42	AC	C1	prospective	yes	78
Chitty et al[201]	1994	UK	649	649	12 - 42	FL	C1	prospective	yes	78
Chitty et al[237]	1994	UK	594	594	12 - 42	BPD, HC	C1	prospective	yes	78

Author	Year	Country	Number of women	Number of scans	Weeks	Measurement	Design	Data collection	Scans for research purposes only	Quality score (%)
De La Vega et al[238]	2008	Puerto Rico	548	548	14 - 38	BPD, HC, AC, FL	C1	unclear	no	13
Deter et al[239]	1982	USA	252	252	13 - 41	HC, AC	C1	unclear	unclear	35
Deter et al[240]	1982	USA	20	unclear	12 - 40	BPD, HC, AC	LL	unclear	unclear	33
Deter et al[241]	1987	USA	20	unclear	15 - 40	FL	LL	unclear	unclear	33
Di Battista et al[242]	2000	Italy	238	1237 - 1539	12 - 40	BPD, HC, AC, FL	LL	unclear	unclear	42
Dubiel et al[243]	2008	Poland	959	959	20 - 42	BPD, HC, AC, FL	CS	prospective	unclear	43
Elejalde et al[244]	1986	USA	1068	2409	10 - 40	BPD, FL	CM	unclear	no	22
Eriksen et al[245]	1985	Denmark	41	493	13 - 40	BPD	LL	prospective	unclear	42
Exacoustos et al[246]	1991	Italy	2317	2317	13 - 40	FL	C1	unclear	no	48
Fescina et al[247]	1982	-	30	692 - 722	13 - 40	BPD, HC, AC	LC	unclear	unclear	25
Flamme[248]	1972	Belgium	1394	4170	20 - 42	BPD	CM	unclear	unclear	0
Gallivan et al[249]	1993	UK	67	434	20 - 40	AC	LL	prospective	yes	58
Guihard-Costa et al[250]	1995	France	3433	3974 - 4248	8 - 41	BPD, FL	CM	retrospective	unclear	30
Hadlock et al[251]	1982	USA	400	400	15 - 41	HC	C1	unclear	unclear	43
Hadlock et al[254]	1982	USA	338	338	12 - 40	FL	CS	unclear	no	35
Hadlock et al[254]	1982	USA	400	400	15 - 41	AC	C1	unclear	unclear	39
Hadlock et al[240]	1982	USA	533	533	12 - 40	BPD	C1	unclear	unclear	43
Hoffbauer et al[255]	1979	Germany	430	800	12 - 40	BPD, HC, AC	CM	unclear	unclear	9
Issel et al[257]	1975	Germany	unclear	5400	12 - 43	BPD	CS	unclear	unclear	9

Author	Year	Country	Number of women	Number of scans	Weeks	Measurement	Design	Data collection	Scans for research purposes only	Quality score (%)
Issel et al[256]	1985	Germany	unclear	1000	17 - 41	FL	CS	retrospective	unclear	9
Jeanty et al[261]	1981	Belgium	450	unclear	15 - 40	FL	CS	unclear	unclear	17
Jeanty et al[258]	1984	Belgium	45	695	12 - 40	AC	LL	prospective	yes	67
Jeanty et al[260]	1984	Belgium	45	696	10 - 40	BPD, HC	LL	prospective	yes	63
Jeanty et al[259]	1984	Belgium	45	646	12 - 40	FL	LL	prospective	yes	63
Johnsen et al[262]	2006	Norway	650	2489 - 2589	10 - 42	BPD, HC, AC, FL	LL	prospective	yes	67
Jung et al[263]	2007	South Korea	10455	10455	12 - 40	BPD, HC, AC, FL	C1	retrospective	no	61
Kurmanavicius et al[264]	1999	Switzerland	6557	5462 - 6217	12 - 42	BPD, HC	C1	retrospective	no	65
Kurmanavicius et al[265]	1999	Switzerland	6557	5807 - 5860	12 - 42	AC, FL	C1	retrospective	no	65
Lai et al[266]	1995	Singapore	6374	6017 - 6131	14 - 41	BPD, HC, AC, FL	C1	retrospective	no	61
Larsen et al[267]	1990	Denmark	35	unclear	14 - 40	BPD, AC, FL	LC	unclear	unclear	42
Lei et al[268]	1998	China	5496	5496	16 - 40	BPD, HC, AC, FL	C1	prospective	no	22
Lessoway et al[269]	1998	Canada	1396	747 - 790	11 - 42	BPD, HC, AC, FL	C1	prospective	no	43
Leung et al[270]	2008	China	709	679 - 708	12 - 40	BPD, HC, AC, FL	C1	prospective	yes	70
Levi et al[272]	1973	Belgium	1011	3032	15 - 43	BPD	CM	unclear	unclear	17
Levi et al[271]	1975	Belgium	unclear	767	19 - 43	BPD, HC	CS	unclear	unclear	9
Lu et al[273]	2008	Taiwan	unclear	50131	14 - 41	AC	CS	retrospective	no	35
Mathai et al[274]	1995	India	120	477 - 498	20 - 40	BPD, AC, FL	LC	prospective	yes	17
Merz et al[275]	1987	Germany	530	530	13 - 42	FL	C1	prospective	unclear	26

Author	Year	Country	Number of women	Number of scans	Weeks	Measurement	Design	Data collection	Scans for research purposes only	Quality score (%)
Munjanja et al[276]	1988	Zimbabwe	190	857 - 1233	12 - 40	BPD, HC	LC	unclear	unclear	13
Munoz et al[277]	1986	South Africa	1842	1842	15 - 40	BPD	C1	unclear	unclear	52
Nasrat et al[11]	2005	Saudi Arabia	unclear	unclear	14 - 40	BPD, HC, AC, FL	C1	retrospective	no	22
Neufeld et al[278]	2004	Guatemala	319	684	14 - 38	BPD, HC, AC, FL	LC	prospective	yes	33
O'Brien et al[279]	1981	USA	411	1016	14 - 40	FL	CM	unclear	unclear	22
Okonofua et al[280]	1988	Nigeria	200	219	20 - 40	AC	CM	unclear	no	17
Okupe et al[281]	1984	Nigeria	552	1104	12 - 40	BPD	CM	unclear	unclear	17
Paladini et al[282]	2005	Italy	626	623 - 625	16 - 40	BPD, HC, AC, FL	C1	prospective	yes	65
Pang et al[283]	2003	China	500	2349	24 - 40	BPD, HC, AC, FL	LL	prospective	yes	50
Persson et al[284]	1978	Sweden	93	707	15 - 42	BPD	LC	unclear	yes	46
Persson et al[285]	1986	Sweden	19	167	11 - 39	BPD, FL	LC	prospective	yes	42
Pineau et al[286]	2003	France	1336	1336	12 - 38	BPD, HC	C1	retrospective	no	39
Queenan et al[287]	1976	USA	468	738	17 - 43	BPD	CM	unclear	unclear	13
Saksiriwuttho et al[288]	2007	Thailand	628	628	14 - 41	BPD, HC, AC, FL	C1	prospective	yes	35
Salomon et al[72]	2006	France	unclear	19647	15 - 40	BPD, HC, AC, FL	CS	unclear	no	48
Schluter et al[289]	2004	Australia	17660	15871 - 20520	11 - 41	BPD, HC, AC, FL	C1	unclear	no	61
Shohat et al[290]	2001	Israel	1143	1143	13 - 41	BPD, FL	C1	prospective	no	22
Siwadune et al[294]	2000	Thailand	613	613	12 - 41	BPD	C1	prospective	yes	65
Smulian et al[292]	2001	USA	10070	10070	11 - 42	AC	CS	retrospective	no	43

Author	Year	Country	Number of women	Number of scans	Weeks	Measurement	Design	Data collection	Scans for research purposes only	Quality score (%)
Snijders et al[293]	1994	UK	1040	1040	14 - 40	BPD, HC, AC, FL	C1	retrospective	no	61
Sunsaneevithayakul et al[294]	2000	Thailand	615	615	12 - 41	AC	C1	prospective	yes	70
Tamura et al[295]	1980	USA	200	536	18 - 41	AC	CM	unclear	unclear	9
Thame et al[296]	2003	Jamaica	499	2574	13 - 37	BPD, HC, AC, FL	LC	prospective	unclear	42
Titapant et al[297]	2000	Thailand	608	608	12 - 41	FL	C1	prospective	yes	70
Todros et al[298]	1987	Italy	1426	603 - 1283	12 - 42	BPD, HC, AC	C1	unclear	unclear	30
Verburg et al[299]	2008	Netherlands	8313	20277-22271	10 - 40	BPD, HC, AC, FL	LL	prospective	yes	75
Warda et al[300]	1985	USA	254	254	13 - 39	FL	C1	unclear	no	57
Watson et al[301]	1986	Australia	202	657	12 - 40	BPD	CM	unclear	yes	26
Westerway et al[302]	2000	Australia	3800	unclear	11 - 40	BPD, HC, AC, FL	CM	unclear	no	22
Wladimiroff et al[303]	1978	Netherlands	303	303	24 - 41	BPD	C1	unclear	no	17

In only 20 out of 83 studies (24%) the ultrasound data were collected prospectively and explicitly for research purposes, whereas in 13 studies (16%) a retrospective analysis of an existing database was performed. In all the remainder it was unclear whether a prospective or retrospective design was used; or whether the ultrasound examinations were performed for research purposes or as part of routine clinical care.

The frequencies of “low risk of bias” in each of the three groups of methodological criteria are presented in Figure 45 - 47. Highest risk of bias was noted in the following fields: “Inclusion / exclusion criteria”, where none of the studies defined a rigorous set of antenatal or fetal conditions that should be excluded from analysis in order to ensure a normal pregnancy outcome (Figure 45, item 1.6); “Ultrasound quality control measures”, where no study demonstrated a comprehensive quality assurance strategy (Figure 47, item 3.6); and “Sample size calculation” which was apparent in only 6 studies (Figure 45, item 1.7).

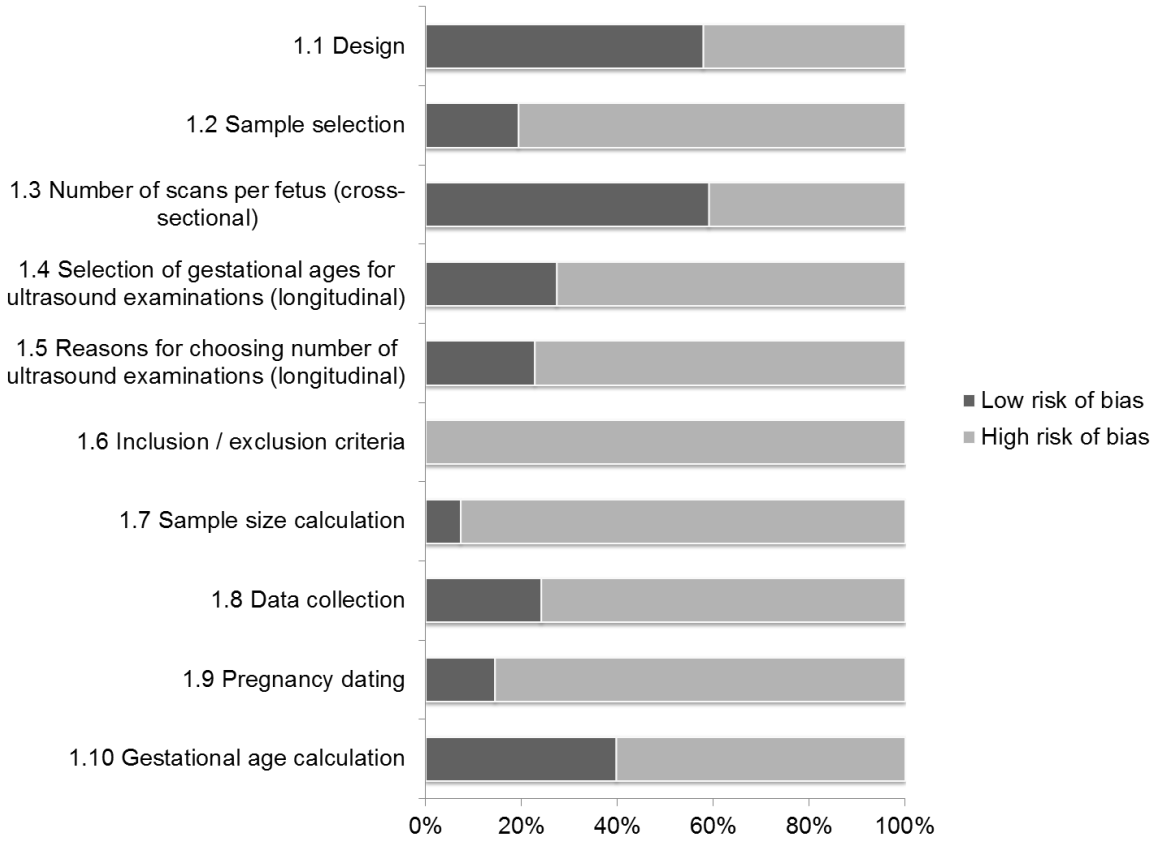


Figure 45 Methodological quality of included studies: Study design

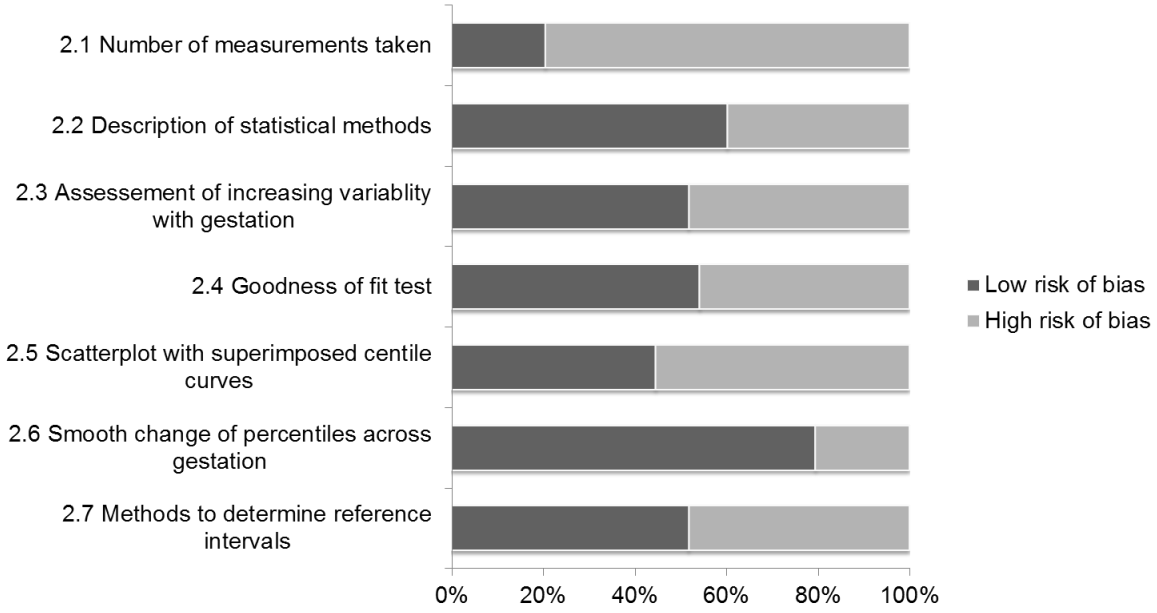


Figure 46 Methodological quality of included studies: Statistical methods

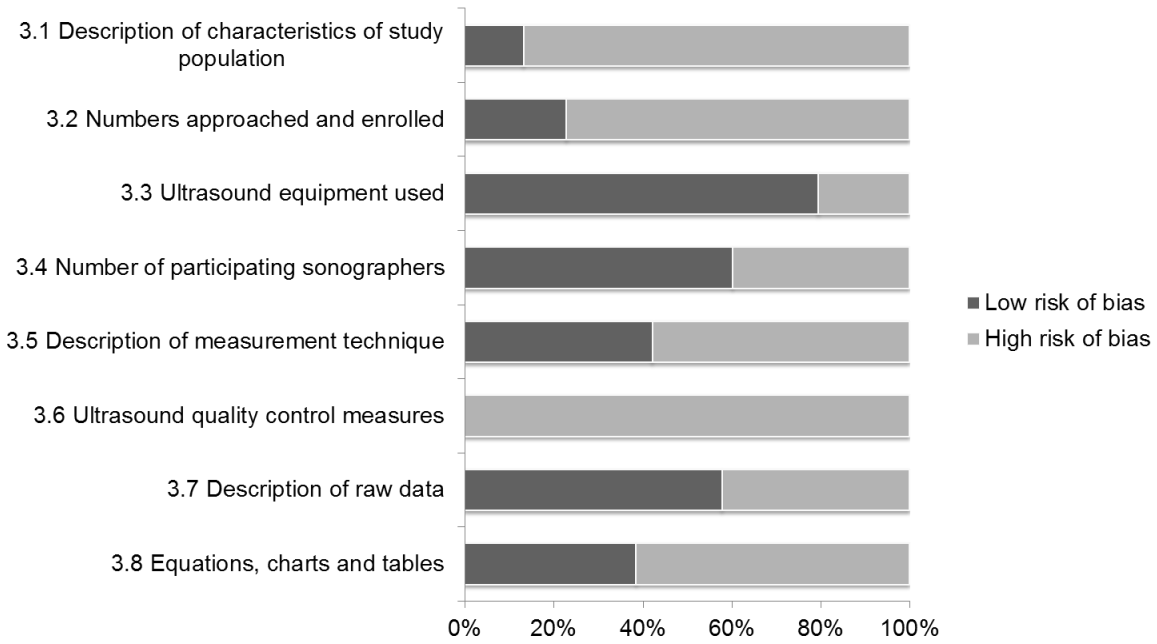


Figure 47 Methodological quality of included studies: Reporting methods

Although some individual criteria of participant selection were used in different studies such as shown in Figure 48, there was no study which systematically used all of them. Conversely, in 22 studies (27%) inappropriate exclusions were applied, such as removing

from final analysis cases on the outer percentiles of the ultrasound measurement or birthweight.

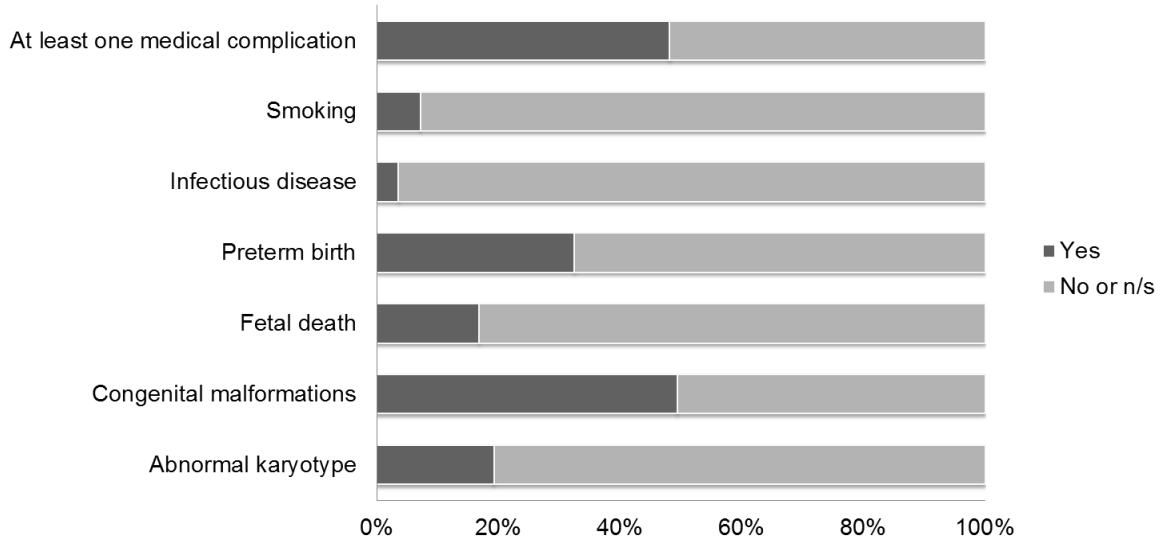


Figure 48 Exclusion criteria used in included studies

Table 22 shows different methods used for gestational age estimation. Last menstrual period (LMP) used in isolation remains the most popular method. Only 12 studies (14%) used a dating method considered to be at low risk of bias, namely either CRL alone or LMP confirmed by CRL.

Table 22 Dating methods (LMP: last menstrual period; US: ultrasound; CRL: crown rump length)

Type of dating	Number of studies (%)
LMP only	37 (45%)
LMP confirmed by US parameter (non-CRL)	21 (25%)
LMP confirmed by CRL	9 (11%)
Other	6 (7%)
Not stated	5 (6%)
CRL only	3 (4%)
US parameter only (non-CRL)	2 (2%)

None of the studies in this review used a comprehensive ultrasound quality control strategy incorporating the items in Figure 49. In approximately half of the studies (40 studies) ultrasound examinations were performed by multiple sonographers, yet an exercise to standardize participating sonographers was reported only in 4 studies. No study reported the use of an image scoring method for the purpose of ultrasound quality assurance; only in 4 studies it was ensured that sonographers were blinded to the actual measurement recorded during the examination.

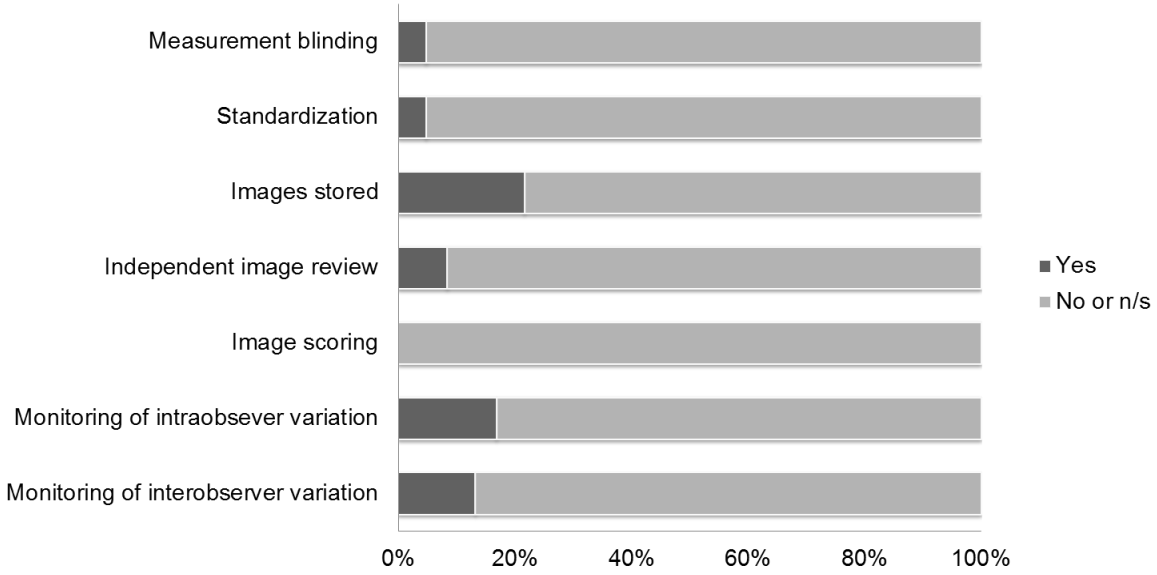


Figure 49 Ultrasound quality assurance measures in included studies

Results from individual studies were reported in the form of tables, equations or charts as demonstrated in Figure 50. Although tables of median values (68 studies) and tables of percentile ranges (65 studies) were a common method of presentation, half of the times these contained the raw un-modelled data; fitted median values following analytical modelling were presented in 35 studies and fitted percentiles in 33 studies. An equation for the median was reported in 50 out of 83 studies whereas the standard deviation was mathematically expressed in 32 out of 83 studies, either as a fixed number or as a function of gestation. Printed charts of the median and percentile curves were seen in the vast majority of the publications.

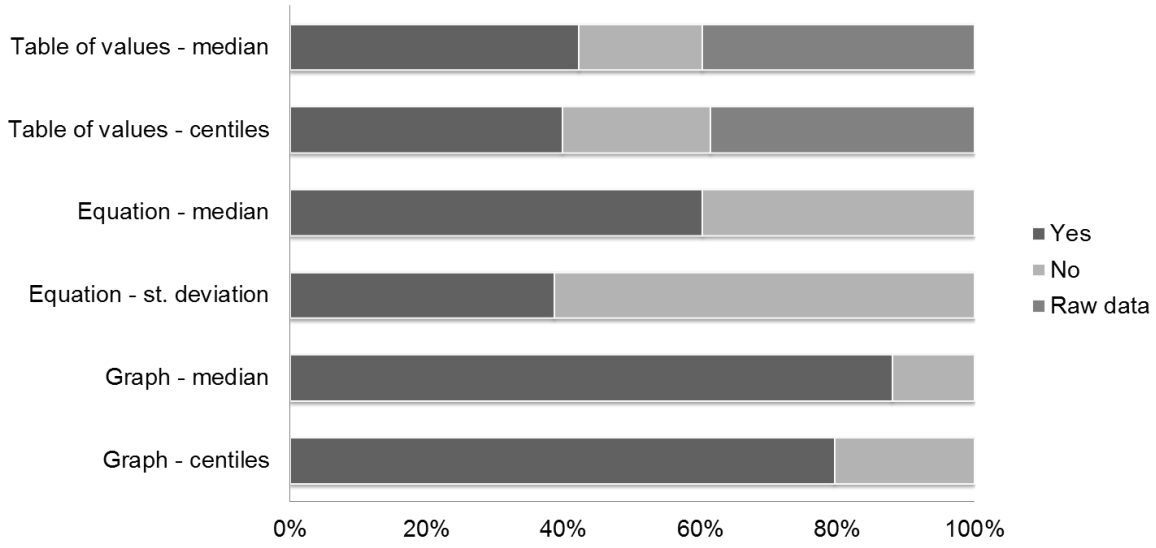


Figure 50 Use of presentation methods in included studies

On univariate regression analysis, positive predictors of quality score were the year of publication ($p < 0.001$), sample size of participating women ($p = 0.04$) and teaching (as opposed to non-teaching) hospital status ($p = 0.003$). On multiple regression analysis however the effect of sample size and hospital type became non-significant, whereas only the year of publication persisted as a significant predictor of quality score with a co-efficient of determination $R^2 = 0.24$.

8.1.4 Discussion

This review has revealed substantial heterogeneity of methodology used in ultrasound studies of fetal biometry. A predefined quality scoring sheet was used in the assessment of included studies. This checklist is not intended to commend or discard studies but rather to be used as a consensus guideline in order to improve consistency in fetal growth research.

This review has several strengths. In the literature search there were no restrictions by year of publication, as some of the older ultrasound charts may still be used in current clinical practice. The quality criteria were based on the best available evidence and were agreed independently of the reviewers who performed the data abstraction. The use of a

quality score in percentage form allowed an objective rather than empirical assessment of quality and also enabled regression analyses in order to identify temporal or other trends. However there are also some limitations. English language restriction was imposed and it is possible that studies of normative fetal biometry from non-English speaking countries may have been missed. This was imposed for practical reasons; unlike systematic reviews of treatment effect, where it is imperative that all evidence is found, our aim was to assess the methodological quality of studies on fetal size. Another limitation was that the reviewers who performed the data abstraction were not blinded to the origin and authors of the included studies. Finally, the older studies in this review were tested against some quality criteria which have only been established in recent years. Although it may seem unfair to judge previously published work by today's standards, it is important for clinicians who choose amongst any of those published ultrasound charts to have an up-to-date assessment of their methodological quality.

Multiple regression analysis demonstrated that quality scores have significantly improved in recent years. This is logical as the present criteria of quality were developed through gradual improvement in both the ultrasound technology and the statistical methods of data analysis. However considerable heterogeneity still persists today: even in the last decade, quality scores of published studies have varied considerably around the average trend. It is possible that such differences in study quality may explain discrepancies in the reported fetal size curves between different populations. Clearly defined and consistent methodology is therefore necessary in order to critically appraise such population differences.

For instance, none of the included studies scored "low risk of bias" for their inclusion and exclusion criteria. This scoring field considered that a study should exclude women with conditions and outcomes associated with pathological fetal growth. The aim of a fetal size chart should be to depict how infants *should* grow under optimal conditions (a "prescriptive" standard) rather than *how they often* grow (a "descriptive" reference)[200]. To achieve this, it is necessary to consider factors that influence growth. A number of such

factors are well established; maternal smoking[209], maternal disease e.g. chronic hypertension or diabetes[210], pregnancy induced hypertension[211], pre-eclampsia[211], abnormal karyotype[210], congenital anomalies[210], pre-term delivery[212] and stillbirth[213]. Some of these conditions were excluded in one or more studies in this review, but no study excluded all of them.

Authors have previously argued against this, claiming that an unselected population ensures a better representation of the underlying population[197]; or that some exclusion criteria, such as maternal smoking for instance, are not reasonable[197, 214]. However, if strong evidence exists for each of them, then all should be excluded. It is unfortunate that the term *super normalization*[214] has acquired a negative connotation in the past.

Conversely, it is arbitrary and illogical to exclude from analysis the outer percentiles of the ultrasound measurement values. For instance, measurements greater than 1.66 or 2 standard deviations from the mean[240, 273] may well represent physiological healthy variability. Similarly, exclusion of cases with birthweight in the outer percentiles has been used by some studies[288, 298] in order to define normal growth. The fundamental concern with this approach is the risk of excluding fetuses which are healthy, but constitutionally small or large for gestational age.

Only a quarter of the studies in this review had a prospective design where ultrasound examinations were performed for research purposes only. This is an important point. Most clinical ultrasound services now routinely collect information in computerized databases. Retrospective analysis of such databases is a practical solution to generating a large sample size; however the ability of the researchers to address potential confounders is curtailed. For instance reference curves can be skewed by the fact that a proportion of the examinations are clinically indicated as a result of suspected pathological growth. It is also difficult to retrospectively ascertain maternal or fetal complications, unless a reliable coding system is in place. The alternative is a prospective study. Strictly speaking, this indicates that the study design, participant recruitment and collection of clinical, demographic and ultrasound data are carried out with the objective of creating size charts.

Due to inconsistent terminology and ambiguous description it was often difficult to identify prospective studies from retrospective database analyses in this review.

Accurate estimation of (gestational) age is a fundamental pre-requisite for creating any size chart. It is already recognised that suboptimal dating in older ultrasound studies may have contributed to the observed flattening or reversal of growth at term[304]. Also, in studies up to the mid-1980s, the research objectives often overlapped between assessing fetal size and pregnancy dating; several studies used a biometric parameter like BPD both as a means of dating and as a predictor of growth, in an almost circular fashion[240, 251, 254].

Several different dating strategies were encountered in this review. On certain occasions dating was either vaguely stated or not described at all. There is now robust evidence that early ultrasonographic determination of gestational age is more reliable than LMP alone[305]. It has also been argued that - for clinical management - ultrasonography alone may be marginally superior to LMP confirmed by ultrasound[306], although such differences are small. However, when the objective is chart creation, the use of ultrasound for dating and then again for assessment of fetal size creates a circular argument; for instance, it is possible that a smaller than expected CRL measurement may be indicative of early growth restriction or adverse pregnancy outcome[307] but such an association will be missed if the gestational age is re-calculated on the basis of CRL. Therefore, in the case of studies aiming to create fetal size charts, we believe it is conceptually more appropriate to estimate the date of confinement using an independent method such as the LMP, provided that this is corroborated by CRL measurement. In the methodological assessment we classified as “low risk of bias” studies that used dating either by CRL alone; or by LMP corroborated by CRL. In any case the parameter used for sonographic dating and the gestational window during which this is applied should always be clearly specified, in order to demonstrate that methods are adequately standardized; this also highlights the recommended dating practice for institutions who wish to adopt those fetal size charts.

Considerable variation exists in collection of ultrasound data, the number of sonographers and scan machines used, measurement method definition, standardization and monitoring of ultrasound data quality. We have proposed a comprehensive quality control strategy which includes saving and independently reviewing scan images; the use of an image scoring method; and the assessment of intra and interobserver variability of measurement. Studies with multiple sonographers are preferable as they reflect real clinical practice, provided that such strategies for quality assurance are in place. Explicit description of measurement planes and calliper application conventions are necessary. A formal standardization exercise prior to the start of a multi-sonographer study has been shown to increase data consistency[308]. Blinding of the sonographers to their own measurements is a reasonable effort to remove observer bias.

A consensus has been reached in recent years regarding the appropriate methods for statistical modelling and data presentation[195-197]. Presentation of the raw measurement data only is not clinically useful or informative. Both the median and variance should be modelled as a function of gestational age in a way that accounts for the increasing variability with gestation and provides smooth percentile curves; goodness of fit testing should demonstrate that these curves describe accurately the structure of the raw data[196]. Finally, the results should be presented in the form of tables of fitted percentile values, gestational curve charts and regression equations for both the mean and the standard deviation.

When we assessed measurement differences from those studies that had the highest scores (Table 23), it was noted that for the majority of biometric parameters significant heterogeneity remains: for example, a BPD measurement of 88 mm at 36 weeks is around the 50th centile of one chart[282] while the same fetus would be below the 10th centile by another[299] and the same is seen at various other gestational ages. Differences in AC were similarly wide, with overlap between 10th and 50th centiles from different studies evident[258, 262]. On the other hand, equivalent centiles of HC seem to be very similar between studies, with most differences well within 0.5 standard deviations.

Table 23 Comparison of measurement values amongst highest scoring studies

BPD studies

Author	28 weeks			32 weeks			36 weeks		
	10th centile	50th centile	SD	10th centile	50th centile	SD	10th centile	50th centile	SD
Chitty[237]	69.29	73.35	3.17	78.95	83.32	3.41	86.78	91.45	3.64
Verburg[299]	70.00	73.68	2.87	80.22	84.38	3.24	88.90	93.55	3.63
Leung[270]	68.09	72.20	3.21	78.04	82.33	3.35	85.55	90.02	3.49
Johnsen[262]	69.00	73.00	3.12	78.00	83.00	3.90	86.00	91.00	3.90
Paladini[282]	66.20	70.90	3.67	75.70	80.30	3.59	84.10	88.10	3.12
Kurmanavicious[264]	70.49	74.94	3.47	79.94	84.74	3.74	87.44	92.56	3.99
Siwadune*[291]	65.71	68.32	2.04	74.93	77.91	2.32	81.75	85.15	2.65

HC studies

Author	28 weeks			32 weeks			36 weeks		
	10th centile	50th centile	SD	10th centile	50th centile	SD	10th centile	50th centile	SD
Chitty[237]	249.10	262.50	10.45	282.70	297.30	11.39	309.00	324.80	12.32
Verburg[299]	251.90	262.70	8.42	285.40	297.80	9.67	309.50	323.50	10.92
Leung[270]	246.90	258.76	9.25	281.22	293.54	9.61	305.60	318.38	9.97
Johnsen[262]	246.00	259.00	10.14	279.00	293.00	10.92	304.00	320.00	12.48
Paladini[282]	250.00	263.50	10.53	280.90	296.10	11.86	303.30	320.00	13.03
Kurmanavicious[264]	248.43	262.73	11.15	279.87	295.32	12.05	303.26	319.87	12.96

AC studies

Author	28 weeks			32 weeks			36 weeks		
	10th centile	50th centile	SD	10th centile	50th centile	SD	10th centile	50th centile	SD
Chitty[236]	212.92	230.57	13.77	248.97	269.71	16.18	282.59	306.41	18.58
Verburg[299]	224.30	239.40	11.78	261.10	278.60	13.65	292.10	312.20	15.68
Leung[270]	218.95	233.90	11.66	256.61	273.56	13.22	290.68	309.64	14.79
Sunsaneevithayakul[294]	217.04	232.58	12.12	254.89	273.26	14.33	289.72	309.12	15.13
Johnsen[262]	223.00	240.00	13.26	262.00	282.00	15.60	299.00	321.00	17.16
Jeanty[258]	208.01	225.24	13.44	245.13	262.36	13.44	276.07	293.30	13.44
Paladini[282]	218.20	239.30	16.46	252.30	275.00	17.71	284.30	307.00	17.71
Kurmanavicious[265]	213.21	231.78	14.49	249.60	270.61	16.39	283.33	306.78	18.29

FL studies

Author	28 weeks			32 weeks			36 weeks		
	10th centile	50th centile	SD	10th centile	50th centile	SD	10th centile	50th centile	SD
Chitty[201]	49.25	52.70	2.69	57.43	61.18	2.93	64.14	68.19	3.16
Verburg[299]	49.68	52.47	2.18	57.43	60.46	2.36	63.24	66.52	2.56
Leung[270]	47.22	50.02	2.18	55.20	58.17	2.32	62.33	65.46	2.44
Titapant[297]	46.45	48.95	1.95	54.39	57.10	2.11	60.95	63.89	2.29
Johnsen[262]	48.00	52.00	3.12	56.00	60.00	3.12	63.00	67.00	3.12
Paladini[282]	49.50	52.40	2.26	56.90	60.50	2.81	62.90	67.20	3.35
Kurmanavicious[265]	48.55	52.30	2.93	56.96	60.85	3.03	64.28	68.28	3.12

For each parameter the five highest scoring studies were identified - additional studies were included if more than one study occupied the fifth place; one study by Ashrafunessa et al[227] is not included despite having the third highest quality score (70%), because data for the fitted centiles were not presented meaning we could not derive values for this table. All BPD measurements were outer to outer except in (*) where the outer to inner convention was used; all measurements quoted in mm; all centiles are fitted centiles following modelling; SD = fitted standard deviation.

This may be a reflection of lower variability in HC measurements across different countries due to the properties of HC as a marker of fat free mass; or may simply be due

to the fact that the HC is a parameter that is more consistently measured in different locations. In other words, whether these differences in biometric parameters are due to different measurement methodologies, or due to differences in population characteristics is impossible to establish using this approach. It is only when studies of fetal size of the highest methodological quality are performed uniformly in different populations, that differences in fetal size can be properly appraised or attributed to biological determinants.

8.1.5 Conclusion

This systematic review has demonstrated considerable heterogeneity of design in ultrasound studies of fetal biometry. Using uniform methodology of the highest quality is essential in order to establish whether population differences in fetal measurements are biological or due to differences in measurement. A checklist of recommended design is proposed in order to aid such uniformity.

8.1.6 Contributions

The author designed the review, extracted the data, performed statistical analyses and created all tables and figures. N Roberts, outreach librarian at the Bodleian Health Care Libraries (Oxford) assisted in the literature search. K Talbot (KT) and E Ohuma (EO) assisted the data extraction; KT contributed in drafting the manuscript. A Conde-Agudelo (ACA), J Villar (JV) and A Papageorghiou (ATP) defined the quality criteria a priori. This systematic review has been published as follows:

- Ioannou C, Talbot K, Ohuma E, Sarris I, Villar J, Conde-Agudelo A, Papageorghiou A. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. BJOG. 2012 Aug 13. doi: 10.1111/j.1471-0528.2012.03451.x. [Epub ahead of print].

Material from this systematic review has also been presented by the author under the title “Systematic review of methodology used in ultrasound studies aimed at creating charts of

fetal size" at the 22nd ISUOG World Congress 2012, Copenhagen, Denmark (poster presentation).

8.2 Publications arising from this thesis

8.2.1 Publications in peer-reviewed journals

- Ioannou C, Sarris I, Salomon LJ and Papageorghiou AT. A review of fetal volumetry: the need for standardization and definitions in measurement methodology. Ultrasound Obstet Gynecol. 2011 Dec;38(6):613-9. doi: 10.1002/uog.9074
- Ioannou C, Sarris I, Yaqub MK, Noble JA, Javaid MK and Papageorghiou AT. Surface area measurement using rendered three-dimensional ultrasound imaging: an in-vitro phantom study. Ultrasound Obstet Gynecol. 2011 Oct;38(4):445-9. doi: 10.1002/uog.8984
- Ioannou C, Sarris I, Javaid MK and Papageorghiou AT. Sphenoidal fontanelle area measurement on rendered three-dimensional ultrasound. Prenat Diagn. 2012 Jun;32(6):592-5. doi: 10.1002/pd.3867
- Ioannou C, Javaid MK, Mahon P, Yaqub MK, Harvey NC, Godfrey KM, Noble JA, Cooper C and Papageorghiou AT. The Effect of Maternal Vitamin D Concentration on Fetal Bone. J Clin Endocrinol Metab. 2012 Sep 18. [Epub ahead of print]
- Ioannou C, Talbot K, Ohuma E, Sarris I, Villar J, Conde-Agudelo A, Papageorghiou A. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. BJOG. 2012 Aug 13. doi: 10.1111/j.1471-0528.2012.03451.x. [Epub ahead of print]

8.2.2 Published abstracts from scientific conferences

- Ioannou C, Yaqub M, Noble A, Javaid K and Papageorghiou A. Fetal femur volume measurement using fusion of multiple three-dimensional ultrasound image

sets: a pilot study. Ultrasound Obstet Gynecol. 2010 Oct;36(S1): 80. doi: 10.1002/uog.8014

- Ioannou C, Sarris I, Yaqub MK, Noble JA, Javaid MK and Papageorghiou AT. Surface area measurement using rendered three-dimensional ultrasound imaging: an in-vitro phantom study. Ultrasound Obstet Gynecol. 2010 Oct;36(S1): 21. doi: 10.1002/uog.7831
- Yaqub M, Ioannou C, Papageorghiou A, Javaid MK, Cooper C, Noble JA. Improving boundary definition for 3D ultrasound quantification of fetal femur. Proceedings of Medical Image Understanding and Analysis 2010: 89-94
- Ioannou C, Sarris I, Yaqub M, Napolitano R, Nicholl H, Noble JA, Javaid MK and Papageorghiou AT. A simple method for calculation of the fetal femur volume in 3D ultrasound. Ultrasound Obstet Gynecol, 40: 289. doi: 10.1002/uog.12180
- Ioannou C, Javaid MK, Mahon P, Yaqub M, Noble JA, Harvey NC, Godfrey KM, Cooper C and Papageorghiou AT. The effect of maternal vitamin D concentration on fetal bone. Ultrasound Obstet Gynecol, 40: 156. doi: 10.1002/uog.11715
- Ioannou C, Talbot K, Ohuma E, Sarris I, Villar J, Conde-Agudelo A. and Papageorghiou AT. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. Ultrasound Obstet Gynecol, 40: 287. doi: 10.1002/uog.12174

9 References

1. Reddy UM, Filly RA,Copel JA. Prenatal imaging: ultrasonography and magnetic resonance imaging. *Obstet Gynecol* 2008; **112**: 145-157.

2. Houston LE, Odibo AO,Macones GA. The safety of obstetrical ultrasound: a review. *Prenat Diagn* 2009; **29**: 1204-1212.

3. Abramowicz JS, Kossoff G, Marsal K, et al. Safety Statement, 2000 (reconfirmed 2003). International Society of Ultrasound in Obstetrics and Gynecology (ISUOG). *Ultrasound Obstet Gynecol* 2003; **21**: 100.

4. AIUM. *Medical Ultrasound Safety*. American Institute of Ultrasound in Medicine 2009.

5. BMUS. *Guidelines for the safe use of diagnostic ultrasound equipment*. British Medical Ultrasound Society, 2009.

6. Torloni MR, Vedmedovska N, Merialdi M, et al. Safety of ultrasonography in pregnancy: WHO systematic review of the literature and meta-analysis. *Ultrasound Obstet Gynecol* 2009; **33**: 599-608.

7. Kot BC, Sin DM,Ying M. Evaluation of the accuracy and reliability of two 3-dimensional sonography methods in volume measurement of small structures: an in vitro phantom study. *J Clin Ultrasound* 2009; **37**: 82-88.

8. MacGillivray TJ, Ross E, Simpson HA, et al. 3D freehand ultrasound for in vivo determination of human skeletal muscle volume. *Ultrasound Med Biol* 2009; **35**: 928-935.

9. Sugeng L, Weinert L, Thiele K, et al. Real-time three-dimensional echocardiography using a novel matrix array transducer. *Echocardiography* 2003; **20**: 623-635.

10. Lapaire O, Alder J, Peukert R, et al. Two- versus three-dimensional ultrasound in the second and third trimester of pregnancy: impact on recognition and maternal-fetal bonding. A prospective pilot study. *Arch Gynecol Obstet* 2007; **276**: 475-479.

11. Nasrat H,Bondagji NS. Ultrasound biometry of Arabian fetuses. *Int J Gynaecol Obstet* 2005; **88**: 173-178.

12. Rustico MA, Mastromatteo C, Grigio M, et al. Two-dimensional vs. two- plus four-dimensional ultrasound in pregnancy and the effect on maternal emotional status: a randomized study. *Ultrasound Obstet Gynecol* 2005; **25**: 468-472.

13. Kurjak A, Miskovic B, Andonotopo W, et al. How useful is 3D and 4D ultrasound in perinatal medicine? *J Perinat Med* 2007; **35**: 10-27.

14. Falcon O, Peralta CF, Cavoretto P, et al. Fetal trunk and head volume measured by three-dimensional ultrasound at 11 + 0 to 13 + 6 weeks of gestation in chromosomally normal pregnancies. *Ultrasound Obstet Gynecol* 2005; **26**: 263-266.

15. Dyson RL, Pretorius DH, Budorick NE, et al. Three-dimensional ultrasound in the evaluation of fetal anomalies. *Ultrasound Obstet Gynecol* 2000; **16**: 321-328.

16. Yagel S, Cohen SM, Shapiro I, et al. 3D and 4D ultrasound in fetal cardiac scanning: a new look at the fetal heart. *Ultrasound Obstet Gynecol* 2007; **29**: 81-95.

17. Chaoui R,Heling KS. Three-dimensional ultrasound in prenatal diagnosis. *Curr Opin Obstet Gynecol* 2006; **18**: 192-202.

18. Benacerraf BR, Shipp TD,Bromley B. Three-dimensional US of the fetus: volume imaging. *Radiology* 2006; **238**: 988-996.

19. Bhaduri M, Fong K, Toi A, et al. Fetal anatomic survey using three-dimensional ultrasound in conjunction with first-trimester nuchal translucency screening. *Prenat Diagn*; **30**: 267-273.

20. Chan LW, Fung TY, Leung TY, et al. Volumetric (3D) imaging reduces inter- and intraobserver variation of fetal biometry measurements. *Ultrasound Obstet Gynecol* 2009; **33**: 447-452.

21. Yang F, Leung KY, Lee YP, et al. Fetal biometry by an inexperienced operator using two- and three-dimensional ultrasound. *Ultrasound Obstet Gynecol*; **35**: 566-571.

22. Borenstein M, Persico N, Kaihura C, et al. Frontomaxillary facial angle in chromosomally normal fetuses at 11 + 0 to 13 + 6 weeks. *Ultrasound Obstet Gynecol* 2007; **30**: 737-741.

23. Lee W, Blanckaert K, Bronsteen RA, et al. Fetal iliac angle measurements by three-dimensional sonography. *Ultrasound Obstet Gynecol* 2001; **18**: 150-154.

24. Perni SC, Chervenak FA, Kalish RB, et al. Intraobserver and interobserver reproducibility of fetal biometry. *Ultrasound Obstet Gynecol* 2004; **24**: 654-658.

25. Chang CH, Tsai PY, Yu CH, et al. Prenatal detection of fetal growth restriction by fetal femur volume: efficacy assessment using three-dimensional ultrasound. *Ultrasound Med Biol* 2007; **33**: 335-341.

26. Correa FF, Lara C, Bellver J, et al. Examination of the fetal brain by transabdominal three-dimensional ultrasound: potential for routine neurosonographic studies. *Ultrasound Obstet Gynecol* 2006; **27**: 503-508.

27. Pilu G, Segata M, Ghi T, et al. Diagnosis of midline anomalies of the fetal brain with the three-dimensional median view. *Ultrasound Obstet Gynecol* 2006; **27**: 522-529.

28. De Jong-Pleij EA, Ribbert LS, Tromp E, et al. Three-dimensional multiplanar ultrasound is a valuable tool in the study of the fetal profile in the second trimester of pregnancy. *Ultrasound Obstet Gynecol*; **35**: 195-200.

29. Roelfsema NM, Grijseels EW, Hop WC, et al. Three-dimensional sonography of prenatal skull base development. *Ultrasound Obstet Gynecol* 2007; **29**: 372-377.

30. Roelfsema NM, Hop WC,Wladimiroff JW. Three-dimensional sonographic determination of normal fetal mandibular and maxillary size during the second half of pregnancy. *Ultrasound Obstet Gynecol* 2006; **28**: 950-957.

31. Rochelson B, Vohra N, Krantz D, et al. Geometric morphometric analysis of shape outlines of the normal and abnormal fetal skull using three-dimensional sonographic multiplanar display. *Ultrasound Obstet Gynecol* 2006; **27**: 167-172.

32. Maarse W, Berge SJ, Pistorius L, et al. Diagnostic accuracy of transabdominal ultrasound in detecting prenatal cleft lip and palate: a systematic review. *Ultrasound Obstet Gynecol*; **35**: 495-502.

33. Ten PM, Pedregosa JP, Santacruz B, et al. Three-dimensional ultrasound diagnosis of cleft palate: 'reverse face', 'flipped face' or 'oblique face'--which method is best? *Ultrasound Obstet Gynecol* 2009; **33**: 399-406.

34. Ginath S, Debby A,Malinger G. Demonstration of cranial sutures and fontanelles at 15 to 16 weeks of gestation: a comparison between two-dimensional and three-dimensional ultrasonography. *Prenat Diagn* 2004; **24**: 812-815.

35. Dikkeboom CM, Roelfsema NM, Van Adrichem LN, et al. The role of three-dimensional ultrasound in visualizing the fetal cranial sutures and fontanels during the second half of pregnancy. *Ultrasound Obstet Gynecol* 2004; **24**: 412-416.

36. Faro C, Benoit B, Wegrzyn P, et al. Three-dimensional sonographic description of the fetal frontal bones and metopic suture. *Ultrasound Obstet Gynecol* 2005; **26**: 618-621.

37. Chaoui R, Levaillant JM, Benoit B, et al. Three-dimensional sonographic description of abnormal metopic suture in second- and third-trimester fetuses. *Ultrasound Obstet Gynecol* 2005; **26**: 761-764.

38. Benacerraf BR, Spiro R,Mitchell AG. Using three-dimensional ultrasound to detect craniosynostosis in a fetus with Pfeiffer syndrome. *Ultrasound Obstet Gynecol* 2000; **16**: 391-394.

39. Krakow D, Santulli T,Platt LD. Use of three-dimensional ultrasonography in differentiating craniosynostosis from severe fetal molding. *J Ultrasound Med* 2001; **20**: 427-431.

40. Paladini D, Vassallo M, Sglavo G, et al. Normal and abnormal development of the fetal anterior fontanelle: a three-dimensional ultrasound study. *Ultrasound Obstet Gynecol* 2008; **32**: 755-761.

41. Paladini D, Sglavo G, Penner I, et al. Fetuses with Down syndrome have an enlarged anterior fontanelle in the second trimester of pregnancy. *Ultrasound Obstet Gynecol* 2007; **30**: 824-829.

42. Raine-Fenning NJ, Clewes JS, Kendall NR, et al. The interobserver reliability and validity of volume calculation from three-dimensional ultrasound datasets in the in vitro setting. *Ultrasound Obstet Gynecol* 2003; **21**: 283-291.

43. Riccabona M, Nelson TR,Pretorius DH. Three-dimensional ultrasound: accuracy of distance and volume measurements. *Ultrasound Obstet Gynecol* 1996; **7**: 429-434.

44. Riccabona M, Nelson TR, Pretorius DH, et al. In vivo three-dimensional sonographic measurement of organ volume: validation in the urinary bladder. *J Ultrasound Med* 1996; **15**: 627-632.

45. Rousian M, Koning AH, van Oppenraaij RH, et al. An innovative virtual reality technique for automated human embryonic volume measurements. *Hum Reprod*; **25**: 2210-2216.

46. Sur SD, Jayaprakasan K, Jones NW, et al. A novel technique for the semi-automated measurement of embryo volume: an intraobserver reliability study. *Ultrasound Med Biol* 2010; **36**: 719-725.

47. Pohls UG,Rempen A. Fetal lung volumetry by three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 1998; **11**: 6-12.

48. Lee A, Kratochwil A, Stumpflen I, et al. Fetal lung volume determination by three-dimensional ultrasonography. *Am J Obstet Gynecol* 1996; **175**: 588-592.

49. Aviram R, Shpan DK, Markovitch O, et al. Three-dimensional first trimester fetal volumetry: comparison with crown rump length. *Early Hum Dev* 2004; **80**: 1-5.

50. Chang CH, Yu CH, Chang FM, et al. The assessment of normal fetal brain volume by 3-D ultrasound. *Ultrasound Med Biol* 2003; **29**: 1267-1272.

51. Roelfsema NM, Hop WC, Boito SM, et al. Three-dimensional sonographic measurement of normal fetal brain volume during the second half of pregnancy. *Am J Obstet Gynecol* 2004; **190**: 275-280.

52. Rutten MJ, Pistorius LR, Mulder EJ, et al. Fetal cerebellar volume and symmetry on 3-d ultrasound: volume measurement with multiplanar and vocal techniques. *Ultrasound Med Biol* 2009; **35**: 1284-1289.

53. Werneck Britto IS, de Silva Bussamra LC, Araujo Junior E, et al. Reference range of fetal lung volume by 3D-ultrasonography using the rotational method (VOCAL). *J Perinat Med* 2009; **37**: 161-167.

54. Gerards FA, Engels MA, Twisk JW, et al. Normal fetal lung volume measured with three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2006; **27**: 134-144.

55. Chang CH, Yu CH, Chang FM, et al. The assessment of normal fetal liver volume by three-dimensional ultrasound. *Ultrasound Med Biol* 2003; **29**: 1123-1129.

56. Chang CH, Tsai PY, Yu CH, et al. Predicting fetal growth restriction with renal volume using 3-D ultrasound: efficacy evaluation. *Ultrasound Med Biol* 2008; **34**: 533-537.

57. Chang CH, Yu CH, Chang FM, et al. Assessment of normal fetal upper arm volume by three-dimensional ultrasound. *Ultrasound Med Biol* 2002; **28**: 859-863.

58. Chang CH, Yu CH, Chang FM, et al. Three-dimensional ultrasound in the assessment of normal fetal thigh volume. *Ultrasound Med Biol* 2003; **29**: 361-366.

59. Chang CH, Yu CH, Ko HC, et al. Prenatal assessment of normal fetal humerus volume by three-dimensional ultrasound. *Ultrasound Med Biol* 2003; **29**: 1675-1680.

60. Wallny T, Schild RL, Fimmers R, et al. Three-dimensional sonographic evaluation of the fetal lumbar spinal canal. *J Anat* 2002; **200**: 439-443.

61. Blaas HG, Eik-Nes SH, Berg S, et al. In-vivo three-dimensional ultrasound reconstructions of embryos and early fetuses. *Lancet* 1998; **352**: 1182-1186.

62. Schuchter K, Metzenbauer M, Hafner E, et al. Uterine artery Doppler and placental volume in the first trimester in the prediction of pregnancy complications. *Ultrasound Obstet Gynecol* 2001; **18**: 590-592.

63. Hafner E, Philipp T, Schuchter K, et al. Second-trimester measurements of placental volume by three-dimensional ultrasound to predict small-for-gestational-age infants. *Ultrasound Obstet Gynecol* 1998; **12**: 97-102.

64. Chang FM, Liang RI, Ko HC, et al. Three-dimensional ultrasound-assessed fetal thigh volumetry in predicting birth weight. *Obstet Gynecol* 1997; **90**: 331-339.

65. Lee W, Balasubramaniam M, Deter RL, et al. New fetal weight estimation models using fractional limb volume. *Ultrasound Obstet Gynecol* 2009; **34**: 556-565.

66. Jani JC, Cannie M, Peralta CF, et al. Lung volumes in fetuses with congenital diaphragmatic hernia: comparison of 3D US and MR imaging assessments. *Radiology* 2007; **244**: 575-582.

67. Gerards FA, Twisk JW, Fetter WP, et al. Predicting pulmonary hypoplasia with 2- or 3-dimensional ultrasonography in complicated pregnancies. *Am J Obstet Gynecol* 2008; **198**: 140 e141-146.

68. Rizzo G, Capponi A, Cavicchioni O, et al. First trimester uterine Doppler and three-dimensional ultrasound placental volume calculation in predicting pre-eclampsia. *Eur J Obstet Gynecol Reprod Biol* 2008; **138**: 147-151.

69. Hafner E, Metzenbauer M, Hofinger D, et al. Comparison between three-dimensional placental volume at 12 weeks and uterine artery impedance/notching at 22 weeks in screening for pregnancy-induced hypertension, pre-eclampsia and fetal growth restriction in a low-risk population. *Ultrasound Obstet Gynecol* 2006; **27**: 652-657.

70. Chen M, Leung KY, Lee CP, et al. Placental volume measured by three-dimensional ultrasound in the prediction of fetal alpha(0)-thalassemia: a preliminary report. *Ultrasound Obstet Gynecol* 2006; **28**: 166-172.

71. Wegrzyn P, Faro C, Falcon O, et al. Placental volume measured by three-dimensional ultrasound at 11 to 13 + 6 weeks of gestation: relation to chromosomal defects. *Ultrasound Obstet Gynecol* 2005; **26**: 28-32.

72. Salomon LJ, Duyme M, Crequat J, et al. French fetal biometry: reference equations and comparison with other charts. *Ultrasound Obstet Gynecol* 2006; **28**: 193-198.

73. Hsieh YY, Chang CC, Lee CC, et al. Fetal renal volume assessment by three-dimensional ultrasonography. *Am J Obstet Gynecol* 2000; **182**: 377-379.

74. Yu C, Chang C, Chang F, et al. Fetal renal volume in normal gestation: a three-dimensional ultrasound study. *Ultrasound Med Biol* 2000; **26**: 1253-1256.

75. Endres LK,Cohen L. Reliability and validity of three-dimensional fetal brain volumes. *J Ultrasound Med* 2001; **20**: 1265-1269.

76. Hata T, Kuno A, Dai S-Y, et al. Three-dimensional sonographic volume measurement of the fetal cerebellum. *Journal of Medical Ultrasonics* 2007; **34**: 17-21.

77. Chang CH, Chang FM, Yu CH, et al. Assessment of fetal cerebellar volume using three-dimensional ultrasound. *Ultrasound Med Biol* 2000; **26**: 981-988.

78. Araujo Junior E, Guimaraes Filho HA, Pires CR, et al. Validation of fetal cerebellar volume by three-dimensional ultrasonography in Brazilian population. *Arch Gynecol Obstet* 2007; **275**: 5-11.

79. Blaas HG, Taipale P, Torp H, et al. Three-dimensional ultrasound volume calculations of human embryos and young fetuses: a study on the volumetry of compound structures and its reproducibility. *Ultrasound Obstet Gynecol* 2006; **27**: 640-646.

80. Gadelha PS, Da Costa AG, Filho FM, et al. Amniotic fluid volumetry by three-dimensional ultrasonography during the first trimester of pregnancy. *Ultrasound Med Biol* 2006; **32**: 1135-1139.

81. Araujo Junior E, Nardozza LM, Rolo LC, et al. Reference range of embryo volume by 3-D sonography using the XI VOCAL method at 7 to 10 + 6 weeks of pregnancy. *Am J Perinatol*; **27**: 501-505.

82. Rolo LC, Nardozza LM, Araujo Junior E, et al. Measurement of embryo volume at 7-10 weeks' gestation by 3D-sonography. *J Obstet Gynaecol* 2009; **29**: 188-191.

83. Bagratee JS, Regan L, Khullar V, et al. Reference intervals of gestational sac, yolk sac and embryo volumes using three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2009; **34**: 503-509.

84. Martins WP, Nastri CO, Barra DA, et al. Fetal volume and crown-rump length from 7 to 10 weeks of gestational age in singletons and twins. *Eur J Obstet Gynecol Reprod Biol* 2009; **145**: 32-35.

85. Mahon PA, Cooper C, Crozier SR, et al. The use of 3D ultrasound to investigate fetal bone development. *Norsk epidemiologi* 2009; **19**: 45-52.

86. Chang FM, Hsu KF, Ko HC, et al. Three-dimensional ultrasound assessment of fetal liver volume in normal pregnancy: a comparison of reproducibility with two-dimensional ultrasound and a search for a volume constant. *Ultrasound Med Biol* 1997; **23**: 381-389.

87. Kuno A, Hayashi Y, Akiyama M, et al. Three-dimensional sonographic measurement of liver volume in the small-for-gestational-age fetus. *J Ultrasound Med* 2002; **21**: 361-366.

88. Boito SM, Laudy JA, Struijk PC, et al. Three-dimensional US assessment of hepatic volume, head circumference, and abdominal circumference in healthy and growth-restricted fetuses. *Radiology* 2002; **223**: 661-665.

89. Laudy JA, Janssen MM, Struyk PC, et al. Fetal liver volume measurement by three-dimensional ultrasonography: a preliminary study. *Ultrasound Obstet Gynecol* 1998; **12**: 93-96.

90. Dos Santos Rizzi MC, Araujo Junior E, Nardozza LM, et al. Nomogram of fetal liver volume by three-dimensional ultrasonography at 27 to 38 weeks of pregnancy using a new multiplanar technique. *Am J Perinatol* 2010; **27**: 641-648.

91. D'Arcy TJ, Hughes SW, Chiu WS, et al. Estimation of fetal lung volume using enhanced 3-dimensional ultrasound: a new method and first result. *Br J Obstet Gynaecol* 1996; **103**: 1015-1020.

92. Hata T, Kuno A, Dai SY, et al. Three-dimensional sonographic volume measurement of the fetal lung. *J Obstet Gynaecol Res* 2007; **33**: 793-798.

93. Osada H, Iitsuka Y, Masuda K, et al. Application of lung volume measurement by three-dimensional ultrasonography for clinical assessment of fetal lung development. *J Ultrasound Med* 2002; **21**: 841-847.

94. Bahmaie A, Hughes SW, Clark T, et al. Serial fetal lung volume measurement using three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2000; **16**: 154-158.

95. Moeglin D, Talmant C, Duyme M, et al. Fetal lung volumetry using two- and three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2005; **25**: 119-127.

96. Araujo Junior E, Nardozza LM, Pires CR, et al. Comparison of the two-dimensional and multiplanar methods and establishment of a new constant for the measurement of fetal lung volume. *J Matern Fetal Neonatal Med* 2008; **21**: 81-88.

97. Chang CH, Yu CH, Chang FM, et al. Volumetric assessment of normal fetal lungs using three-dimensional ultrasound. *Ultrasound Med Biol* 2003; **29**: 935-942.

98. Britto IS, de Silva Bussamra LC, Araujo Junior E, et al. Fetal lung volume: comparison by 2D- and 3D-sonography in normal fetuses. *Arch Gynecol Obstet* 2009; **280**: 363-368.

99. Peralta CF, Cavoretto P, Csapo B, et al. Lung and heart volumes by three-dimensional ultrasound in normal fetuses at 12-32 weeks' gestation. *Ultrasound Obstet Gynecol* 2006; **27**: 128-133.

100. Ruano R, Benachi A, Joubin L, et al. Three-dimensional ultrasonographic assessment of fetal lung volume as prognostic factor in isolated congenital diaphragmatic hernia. *BJOG* 2004; **111**: 423-429.

101. Laudy JA, Janssen MM, Struyk PC, et al. Three-dimensional ultrasonography of normal fetal lung volume: a preliminary study. *Ultrasound Obstet Gynecol* 1998; **11**: 13-16.

102. Sabogal JC, Becker E, Bega G, et al. Reproducibility of fetal lung volume measurements with 3-dimensional ultrasonography. *J Ultrasound Med* 2004; **23**: 347-352.

103. Tedesco GD, Bussamra LC, Araujo Junior E, et al. Reference range of fetal renal volume by three-dimensional ultrasonography using the VOCAL method. *Fetal Diagn Ther* 2009; **25**: 385-391.

104. Salomon LJ, Bernard JP, Duyme M, et al. Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester. *Ultrasound Obstet Gynecol* 2006; **27**: 34-40.

105. Sairam S, Awadh AM, Cook K, et al. Impact of audit of routine second-trimester cardiac images using a novel image-scoring method. *Ultrasound Obstet Gynecol* 2009; **33**: 545-551.

106. Herman A, Dreazen E, Maymon R, et al. Implementation of nuchal translucency image-scoring method during ongoing audit. *Ultrasound Obstet Gynecol* 1999; **14**: 388-392.

107. McGraw KO,Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; **1**: 30-46.

108. Bland JM,Altman DG. Measurement error. *BMJ* 1996; **313**: 744.

109. Bland JM,Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307-310.

110. Thayyil S, Chitty LS, Robertson NJ, et al. Minimally invasive fetal postmortem examination using magnetic resonance imaging and computerised tomography: current evidence and practical issues. *Prenat Diagn* 2010; **30**: 713-718.

111. Rutherford MA. Magnetic resonance imaging of the fetal brain. *Curr Opin Obstet Gynecol* 2009; **21**: 180-186.

112. Peralta CF, Kazan-Tannus JF, Bunduki V, et al. Evaluation of the agreement between 3-dimensional ultrasonography and magnetic resonance imaging for fetal lung volume measurement. *J Ultrasound Med* 2006; **25**: 461-467.

113. Braillon PM, Buenerd A, Lapillonne A, et al. Skeletal and total body volumes of human fetuses: assessment of reference data by spiral CT. *Pediatric radiology* 2002; **32**: 354-359.

114. Karaplis AC, John PB, Lawrence GR, et al. Embryonic Development of Bone and the Molecular Regulation of Intramembranous and Endochondral Bone Formation. In *Principles of Bone Biology (Second Edition)*. Academic Press: San Diego, 2002, 33-58.

115. Noden DM. Origins and patterning of craniofacial mesenchymal tissues. *J Craniofac Genet Dev Biol Suppl* 1986; **2**: 15-31.

116. Opperman LA. Cranial sutures as intramembranous bone growth sites. *Dev Dyn* 2000; **219**: 472-485.

117. Couly GF, Coltey PM,Le Douarin NM. The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. *Development* 1993; **117**: 409-429.

118. Cohen MM, Jr. Sutural biology and the correlates of craniosynostosis. *Am J Med Genet* 1993; **47**: 581-616.

119. Wilkie AO, Tang Z, Elanko N, et al. Functional haploinsufficiency of the human homeobox gene MSX2 causes defects in skull ossification. *Nat Genet* 2000; **24**: 387-390.

120. Jabs EW, Muller U, Li X, et al. A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. *Cell* 1993; **75**: 443-450.

121. Muenke M, Schell U, Hehr A, et al. A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nat Genet* 1994; **8**: 269-274.

122. Reardon W, Winter RM, Rutland P, et al. Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome. *Nat Genet* 1994; **8**: 98-103.

123. Christ B, Schmidt C, Huang R, et al. Segmentation of the vertebrate body. *Anat Embryol (Berl)* 1998; **197**: 1-8.

124. Ruvinsky I,Gibson-Brown JJ. Genetic and developmental bases of serial homology in vertebrate limb evolution. *Development* 2000; **127**: 5233-5244.

125. Gottlieb AG,Galan HL. Nontraditional sonographic pearls in estimating gestational age. *Semin Perinatol* 2008; **32**: 154-160.

126. Niswander L. Limb mutants: what can they tell us about normal limb development? *Curr Opin Genet Dev* 1997; **7**: 530-536.

127. Stevens DA,Williams GR. Hormone regulation of chondrocyte differentiation and endochondral bone formation. *Mol Cell Endocrinol* 1999; **151**: 195-204.

128. Bellus GA, Hefferon TW, Ortiz de Luna RI, et al. Achondroplasia is defined by recurrent G380R mutations of FGFR3. *Am J Hum Genet* 1995; **56**: 368-373.

129. Tavormina PL, Shiang R, Thompson LM, et al. Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nat Genet* 1995; **9**: 321-328.

130. Bellus GA, McIntosh I, Smith EA, et al. A recurrent mutation in the tyrosine kinase domain of fibroblast growth factor receptor 3 causes hypochondroplasia. *Nat Genet* 1995; **10**: 357-359.

131. Norman AW. Sunlight, season, skin pigmentation, vitamin D, and 25-hydroxyvitamin D: integral components of the vitamin D endocrine system. *Am J Clin Nutr* 1998; **67**: 1108-1110.

132. Armas LA, Hollis BW,Heaney RP. Vitamin D2 is much less effective than vitamin D3 in humans. *J Clin Endocrinol Metab* 2004; **89**: 5387-5391.

133. Ponchon G, Kennan AL,DeLuca HF. "Activation" of vitamin D by the liver. *J Clin Invest* 1969; **48**: 2032-2037.

134. Armbrecht HJ, Hodam TL,Boltz MA. Hormonal regulation of 25-hydroxyvitamin D3-1alpha-hydroxylase and 24-hydroxylase gene transcription in opossum kidney cells. *Arch Biochem Biophys* 2003; **409**: 298-304.

135. Lips P. Vitamin D physiology. *Prog Biophys Mol Biol* 2006; **92**: 4-8.

136. Jimenez-Lara AM,Aranda A. Interaction of vitamin D and retinoid receptors on regulation of gene expression. *Horm Res* 2000; **54**: 301-305.

137. Kimball S, Fuleihan Gel H,Vieth R. Vitamin D: a growing perspective. *Crit Rev Clin Lab Sci* 2008; **45**: 339-414.

138. Brooke OG, Brown IR, Bone CD, et al. Vitamin D supplements in pregnant Asian women: effects on calcium status and fetal growth. *Br Med J* 1980; **280**: 751-754.

139. Carlsen NL, Krasilnikoff PA,Eiken M. Premature cranial synostosis in X-linked hypophosphatemic rickets: possible precipitation by 1-alpha-OH-cholecalciferol intoxication. *Acta Paediatrica Scandinavica* 1984; **73**: 149-154.

140. Stein EM, Laing EM, Hall DB, et al. Serum 25-hydroxyvitamin D concentrations in girls aged 4-8 y living in the southeastern United States. *Am J Clin Nutr* 2006; **83**: 75-81.

141. Ala-Houhala M, Parviainen MT, Pyykko K, et al. Serum 25-hydroxyvitamin D levels in Finnish children aged 2 to 17 years. *Acta Paediatr Scand* 1984; **73**: 232-236.

142. Guillemant J, Cabrol S, Allemandou A, et al. Vitamin D-dependent seasonal variation of PTH in growing male adolescents. *Bone* 1995; **17**: 513-516.

143. Andersen R, Molgaard C, Skovgaard LT, et al. Teenage girls and elderly women living in northern Europe have low winter vitamin D status. *Eur J Clin Nutr* 2005; **59**: 533-541.

144. Harkness LS,Cromer BA. Vitamin D deficiency in adolescent females. *J Adolesc Health* 2005; **37**: 75.

145. Outila TA, Karkkainen MU,Lamberg-Allardt CJ. Vitamin D status affects serum parathyroid hormone concentrations during winter in female adolescents: associations with forearm bone mineral density. *Am J Clin Nutr* 2001; **74**: 206-210.

146. Hill TR, Cotter AA, Mitchell S, et al. Vitamin D status and its determinants in adolescents from the Northern Ireland Young Hearts 2000 cohort. *Br J Nutr* 2008; **99**: 1061-1067.

147. Malabanan A, Veronikis IE,Holick MF. Redefining vitamin D insufficiency. *Lancet* 1998; **351**: 805-806.

148. Bischoff-Ferrari HA, Giovannucci E, Willett WC, et al. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. *Am J Clin Nutr* 2006; **84**: 18-28.

149. Greer FR. 25-Hydroxyvitamin D: functional outcomes in infants and young children. *Am J Clin Nutr* 2008; **88**: 529S-533S.

150. NICE. *Antenatal care: Routine care for the healthy pregnant woman (NICE clinical guideline 62)*. National Collaborating Centre for Women's and Children's Health, 2008.

151. Ward LM, Gaboury I, Ladhani M, et al. Vitamin D-deficiency rickets among children in Canada. *CMAJ* 2007; **177**: 161-166.

152. Tylavsky FA, Ryder KA, Lyytikainen A, et al. Vitamin D, parathyroid hormone, and bone mass in adolescents. *J Nutr* 2005; **135**: 2735S-2738S.

153. Grover SR,Morley R. Vitamin D deficiency in veiled or dark-skinned pregnant women. *Med J Aust* 2001; **175**: 251-252.

154. Henriksen C, Brunvand L, Stoltenberg C, et al. Diet and vitamin D status among pregnant Pakistani women in Oslo. *Eur J Clin Nutr* 1995; **49**: 211-218.

155. Finch PJ, Ang L, Colston KW, et al. Blunted seasonal variation in serum 25-hydroxy vitamin D and increased risk of osteomalacia in vegetarian London Asians. *Eur J Clin Nutr* 1992; **46**: 509-515.

156. Munns C, Zacharin MR, Rodda CP, et al. Prevention and treatment of infant and childhood vitamin D deficiency in Australia and New Zealand: a consensus statement. *Med J Aust* 2006; **185**: 268-272.

157. Ginde AA, Sullivan AF, Mansbach JM, et al. Vitamin D insufficiency in pregnant and nonpregnant women of childbearing age in the United States. *Am J Obstet Gynecol*; **202**: 436 e431-438.

158. Ritchie LD, Fung EB, Halloran BP, et al. A longitudinal study of calcium homeostasis during human pregnancy and lactation and after resumption of menses. *Am J Clin Nutr* 1998; **67**: 693-701.

159. Prentice A. Calcium in pregnancy and lactation. *Annu Rev Nutr* 2000; **20**: 249-272.

160. Gambacciani M, Spinetti A, Gallo R, et al. Ultrasonographic bone characteristics during normal pregnancy: longitudinal and cross-sectional evaluation. *Am J Obstet Gynecol* 1995; **173**: 890-893.

161. Datta S, Alfaham M, Davies DP, et al. Vitamin D deficiency in pregnant women from a non-European ethnic minority population--an interventional study. *BJOG* 2002; **109**: 905-908.

162. Javaid MK, Crozier SR, Harvey NC, et al. Maternal vitamin D status during pregnancy and childhood bone mass at age 9 years: a longitudinal study. *Lancet* 2006; **367**: 36-43.

163. Bowyer L, Catling-Paull C, Diamond T, et al. Vitamin D, PTH and calcium levels in pregnant women and their neonates. *Clin Endocrinol (Oxf)* 2009; **70**: 372-377.

164. Delvin EE, Salle BL, Glorieux FH, et al. Vitamin D supplementation during pregnancy: effect on neonatal calcium homeostasis. *J Pediatr* 1986; **109**: 328-334.

165. Mallet E, Gugi B, Brunelle P, et al. Vitamin D supplementation in pregnancy: a controlled trial of two methods. *Obstet Gynecol* 1986; **68**: 300-304.

166. Vitamin D supplementation: Recommendations for Canadian mothers and infants. *Paediatr Child Health* 2007; **12**: 583-598.

167. Mahon P, Harvey N, Crozier S, et al. Low maternal vitamin D status and fetal bone development: cohort study. *J Bone Miner Res* 2010; **25**: 14-19.

168. Godfrey K, Walker-Bone K, Robinson S, et al. Neonatal bone mass: influence of parental birthweight, maternal smoking, body composition, and activity during pregnancy. *J Bone Miner Res* 2001; **16**: 1694-1703.

169. McGrath JJ, Keeping D, Saha S, et al. Seasonal fluctuations in birth weight and neonatal limb length; does prenatal vitamin D influence neonatal size and shape? *Early Hum Dev* 2005; **81**: 609-618.

170. Murray LJ, O'Reilly DP, Betts N, et al. Season and outdoor ambient temperature: effects on birth weight. *Obstet Gynecol* 2000; **96**: 689-695.

171. Javaid MK,Cooper C. Prenatal and childhood influences on osteoporosis. *Best Pract Res Clin Endocrinol Metab* 2002; **16**: 349-367.

172. Cooper C, Cawley M, Bhalla A, et al. Childhood growth, physical activity, and peak bone mass in women. *J Bone Miner Res* 1995; **10**: 940-947.

173. Cooper C, Fall C, Egger P, et al. Growth in infancy and bone mass in later life. *Ann Rheum Dis* 1997; **56**: 17-21.

174. Cooper C, Eriksson JG, Forsen T, et al. Maternal height, childhood growth and risk of hip fracture in later life: a longitudinal study. *Osteoporos Int* 2001; **12**: 623-629.

175. Marya RK, Saini AS,Jaswal TS. Effect of vitamin D supplementation during pregnancy on the neonatal skeletal growth in the rat. *Ann Nutr Metab* 1991; **35**: 208-212.

176. Congdon P, Horsman A, Kirby PA, et al. Mineral content of the forearms of babies born to Asian and white mothers. *Br Med J (Clin Res Ed)* 1983; **286**: 1233-1235.

177. Marya RK, Rathee S, Lata V, et al. Effects of vitamin D supplementation in pregnancy. *Gynecol Obstet Invest* 1981; **12**: 155-161.

178. Marya RK, Rathee S, Dua V, et al. Effect of vitamin D supplementation during pregnancy on foetal growth. *Indian J Med Res* 1988; **88**: 488-492.

179. Mahomed K,Gulmezoglu AM. Vitamin D supplementation in pregnancy. *Cochrane Database Syst Rev* 2000: CD000228.

180. Brooke OG, Butters F,Wood C. Intrauterine vitamin D nutrition and postnatal growth in Asian infants. *Br Med J (Clin Res Ed)* 1981; **283**: 1024.

181. Cardinal HN, Gill JD,Fenster A. Analysis of geometrical distortion and statistical variance in length, area, and volume in a linearly scanned 3-D ultrasound image. *IEEE Trans Med Imaging* 2000; **19**: 632-651.

182. Shaffer JP. Multiple Hypothesis-Testing. *Annual Review of Psychology* 1995; **46**: 561-584.

183. Deurloo K, Spreeuwenberg M, Rekoert-Hollander M, et al. Reproducibility of 3-dimensional sonographic measurements of fetal and placental volume at gestational ages of 11-18 weeks. *J Clin Ultrasound* 2007; **35**: 125-132.

184. Duin LK, Willekes C, Vossen M, et al. Reproducibility of fetal renal pelvis volume measurement using three-dimensional ultrasound. *Ultrasound Obstet Gynecol* 2008; **31**: 657-661.

185. Ioannou C, Sarris I, Yaqub MK, et al. Surface area measurement using rendered three-dimensional ultrasound imaging: an in-vitro phantom study. *Ultrasound Obstet Gynecol* 2011; **38**: 445-449.

186. Grau V, Becher H,Noble JA. Registration of multiview real-time 3-D echocardiographic sequences. *IEEE Trans Med Imaging* 2007; **26**: 1154-1165.

187. Gooding MJ, Rajpoot K, Mitchell S, et al. Investigation into the fusion of multiple 4-D fetal echocardiography images to improve image quality. *Ultrasound Med Biol* 2010; **36**: 957-966.

188. Yaqub M, Ioannou C, Papageorghiou A, et al. Improving boundary definition for 3D ultrasound quantification of fetal femur. *Proceedings of Medical Image Understanding and Analysis* 2010: 89-94.

189. Hadlock FP, Kent WR, Loyd JL, et al. An evaluation of two methods for measuring fetal head and body circumferences. *J Ultrasound Med* 1982; **1**: 359-360.

190. Martins WP, Raine-Fenning NJ, Leite SP, et al. A standardized measurement technique may improve the reliability of measurements of endometrial thickness and volume. *Ultrasound in Obstetrics & Gynecology* 2011; **38**: 107-115.

191. Raine-Fenning NJ, Campbell BK, Clewes JS, et al. The interobserver reliability of ovarian volume measurement is improved with three-dimensional ultrasound, but dependent upon technique. *Ultrasound in Medicine & Biology* 2003; **29**: 1685-1690.

192. Smulian JC, Ranzini AC, Ananth CV, et al. Comparison of three sonographic circumference measurement techniques to predict birth weight. *Obstet Gynecol* 1999; **93**: 692-696.

193. Harris J,Stocker H. *Handbook of mathematics and computational science*. Springer: New York ; London, 1998, 112.

194. Khan KS,Chien PF. Evaluation of a clinical test. I: assessment of reliability. *BJOG* 2001; **108**: 562-567.

195. Royston P,Altman DG. Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995; **6**: 307-312.

196. Altman DG,Chitty LS. Charts of fetal size: 1. Methodology. *Br J Obstet Gynaecol* 1994; **101**: 29-34.

197. Altman DG,Chitty LS. Design and analysis of studies to derive charts of fetal size. *Ultrasound Obstet Gynecol* 1993; **3**: 378-384.

198. *Physical status : the use and interpretation of anthropometry. Report of a WHO Expert Committee*. WHO, 1995.

199. Rasbash J, Steele F, J. BW, et al. *A User's Guide to MLwiN*. Centre for Multilevel Modelling, University of Bristol, 2009.

200. de Onis M, Garza C, Victora CG, et al. The WHO Multicentre Growth Reference Study: planning, study design, and methodology. *Food Nutr Bull* 2004; **25**: S15-26.

201. Chitty LS, Altman DG, Henderson A, et al. Charts of fetal size: 4. Femur length. *Br J Obstet Gynaecol* 1994; **101**: 132-135.

202. Hadlock FP, Harrist RB, Carpenter RJ, et al. Sonographic estimation of fetal weight. The value of femur length in addition to head and abdomen measurements. *Radiology* 1984; **150**: 535-540.

203. Smith-Bindman R, Hosmer W, Feldstein VA, et al. Second-trimester ultrasound to detect fetuses with Down syndrome: a meta-analysis. *JAMA* 2001; **285**: 1044-1055.

204. Papageorghiou AT, Fratelli N, Leslie K, et al. Outcome of fetuses with antenatally diagnosed short femur. *Ultrasound Obstet Gynecol* 2008; **31**: 507-511.

205. Inskip HM, Godfrey KM, Robinson SM, et al. Cohort profile: The Southampton Women's Survey. *Int J Epidemiol* 2006; **35**: 42-48.

206. Janssen MJ, Wielders JP, Bekker CC, et al. Multicenter comparison study of current methods to measure 25-hydroxyvitamin D in serum. *Steroids* 2012.

207. Hull H, He Q, Thornton J, et al. iDXA, Prodigy, and DPXL dual-energy X-ray absorptiometry whole-body scans: a cross-calibration study. *J Clin Densitom* 2009; **12**: 95-102.

208. Brenner DJ,Hall EJ. Computed tomography--an increasing source of radiation exposure. *N Engl J Med* 2007; **357**: 2277-2284.

209. DiFranza JR,Lew RA. Effect of maternal cigarette smoking on pregnancy complications and sudden infant death syndrome. *J Fam Pract* 1995; **40**: 385-394.

210. Hendrix N,Berghella V. Non-placental causes of intrauterine growth restriction. *Semin Perinatol* 2008; **32**: 161-165.

211. Rasmussen S,Irgens LM. The effects of smoking and hypertensive disorders on fetal growth. *BMC Pregnancy Childbirth* 2006; **6**: 16.

212. Zeitlin J, Ancel PY, Saurel-Cubizolles MJ, et al. The relationship between intrauterine growth restriction and preterm delivery: an empirical approach using data from a European case-control study. *BJOG* 2000; **107**: 750-758.

213. Stacey T, Thompson JM, Mitchell EA, et al. The Auckland Stillbirth study, a case-control study exploring modifiable risk factors for third trimester stillbirth: methods and rationale. *Aust N Z J Obstet Gynaecol* 2011; **51**: 3-8.

214. Rasmussen S. Charts to assess fetal wellbeing. *Ultrasound Obstet Gynecol* 2011; **37**: 2-5.

215. Campbell S. An improved method of fetal cephalometry by ultrasound. *J Obstet Gynaecol Br Commonw* 1968; **75**: 568-576.

216. Bricker L, Neilson JP,Dowswell T. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane Database Syst Rev* 2008: CD001451.

217. *Centre for Maternal and Child Enquiries (CMACE) Perinatal Mortality 2009*: London, 2011.

218. Raman S, Teoh T,Nagaraj S. Growth patterns of the humeral and femur length in a multiethnic population. *Int J Gynaecol Obstet* 1996; **54**: 143-147.

219. Schwarzler P, Bland JM, Holden D, et al. Sex-specific antenatal reference growth charts for uncomplicated singleton pregnancies at 15-40 weeks of gestation. *Ultrasound Obstet Gynecol* 2004; **23**: 23-29.

220. Krampl E, Lees C, Bland JM, et al. Fetal biometry at 4300 m compared to sea level in Peru. *Ultrasound Obstet Gynecol* 2000; **16**: 9-18.

221. Silverwood RJ,Cole TJ. Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. *Ultrasound Obstet Gynecol* 2007; **29**: 6-13.

222. Royston P,Wright EM. How to construct 'normal ranges' for fetal variables. *Ultrasound Obstet Gynecol* 1998; **11**: 30-38.

223. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**: 2008-2012.

224. Aickin DR, Chatfield WR,Rippon PE. Fetal Cephalometry During Normal Pregnancy in a New Zealand Population. *Australian and New Zealand Journal of Obstetrics and Gynaecology* 1976; **16**: 210-212.

225. Al-Meshari AA,Raber H. Fetal Biparietal Diameter in Saudi Arabia. *Annals of Saudi Medicine* 1987; **7**: 227-233.

226. Amoa AB, Wapi J,Klufio CA. Longitudinal fetal biometry of normal pregnant Melanesian Papua New Guineans to construct standards of reference for Papua New Guinea. *P N G Med J* 1993; **36**: 219-227.

227. Ashrafunnessa, Jehan AH, Chowdhury SB, et al. Construction of fetal charts for biparietal diameter, fetal abdominal circumference and femur length in Bangladeshi population. *Bangladesh Med Res Counc Bull* 2003; **29**: 67-77.

228. Ayangade SO,Okonofua FE. Normal growth of the fetal biparietal diameter in an African population. *Int J Gynaecol Obstet* 1986; **24**: 35-42.

229. Beigi A,ZarrinKoub F. Ultrasound assessment of fetal biparietal diameter and femur length during normal pregnancy in Iranian women. *Int J Gynaecol Obstet* 2000; **69**: 237-242.

230. Bergsjo P, Bakke T,Bjerkedal T. Growth of the fetal skull, with special reference to weight-for-dates of the newborn child. *Acta Obstet Gynecol Scand* 1976; **55**: 53-57.

231. Brons JT, van Geijn HP, Bezemer PD, et al. The fetal skeleton; ultrasonographic evaluation of the normal growth. *European Journal of Obstetrics, Gynecology, & Reproductive Biology* 1990; **34**: 21-36.

232. Browne PC, Hamner LH, 3rd,Clark WS. Sonographic fetal growth curves from an indigent population in Atlanta, Georgia: I. Singleton pregnancies. *American Journal of Perinatology* 1992; **9**: 467-476.

233. Campbell S,Newman GB. Growth of the fetal biparietal diameter during normal pregnancy. *J Obstet Gynaecol Br Commonw* 1971; **78**: 513-519.

234. Chan WB,Yeo GS. A comparison of fetal biparietal diameter measurements between local Chinese and Caucasian populations. *Singapore Med J* 1991; **32**: 214-217.

235. Chang CH, Chang FM, Yu CH, et al. Re-analysis of fetal biparietal diameter during gestation in Taiwanese by Altman's method. *Journal of Medical Ultrasound* 1996; **4**: 162-168.

236. Chitty LS, Altman DG, Henderson A, et al. Charts of fetal size: 3. Abdominal measurements. *Br J Obstet Gynaecol* 1994; **101**: 125-131.

237. Chitty LS, Altman DG, Henderson A, et al. Charts of fetal size: 2. Head measurements. *Br J Obstet Gynaecol* 1994; **101**: 35-43.

238. de la Vega A, Ruiz-Febo N,Roberts ZC. Fetal ultrasound biometry: normative charts for a Puerto Rican population. *P R Health Sci J* 2008; **27**: 81-84.

239. Deter RL, Harrist RB, Hadlock FP, et al. Fetal head and abdominal circumferences: II. A critical re-evaluation of the relationship to menstrual age. *J Clin Ultrasound* 1982; **10**: 365-372.

240. Deter RL, Harrist RB, Hadlock FP, et al. Longitudinal studies of fetal growth with the use of dynamic image ultrasonography. *American Journal of Obstetrics & Gynecology* 1982; **143**: 545-554.

241. Deter RL, Rossavik IK, Hill RM, et al. Longitudinal studies of femur growth in normal fetuses. *J Clin Ultrasound* 1987; **15**: 299-305.

242. Di Battista E, Bertino E, Benso L, et al. Longitudinal distance standards of fetal growth. Intrauterine and Infant Longitudinal Growth Study: IILGS. *Acta Obstetricia et Gynecologica Scandinavica* 2000; **79**: 165-173.

243. Dubiel M, Krajewski M, Pietryga M, et al. Fetal biometry between 20-42 weeks of gestation for Polish population. *Ginekol Pol* 2008; **79**: 746-753.

244. Elejalde BR,de Elejalde MM. The prenatal growth of the human body determined by the measurement of bones and organs by ultrasonography. *Am J Med Genet* 1986; **24**: 575-598.

245. Eriksen PS, Secher NJ,Weis-Bentzon M. Normal growth of the fetal biparietal diameter and the abdominal diameter in a longitudinal study. An evaluation of the two parameters in predicting fetal weight. *Acta Obstet Gynecol Scand* 1985; **64**: 65-70.

246. Exacoustos C, Rosati P, Rizzo G, et al. Ultrasound measurements of fetal limb bones. *Ultrasound Obstet Gynecol* 1991; **1**: 325-330.

247. Fescina RH, Ucieda FJ, Cordano MC, et al. Ultrasonic patterns of intrauterine fetal growth in a Latin American country. *Early Hum Dev* 1982; **6**: 239-248.

248. Flamme P. Ultrasonic fetal cephalometry: percentiles curve. *Br Med J* 1972; **3**: 384-385.

249. Gallivan S, Robson SC, Chang TC, et al. An investigation of fetal growth using serial ultrasound data. *Ultrasound Obstet Gynecol* 1993; **3**: 109-114.

250. Guihard-Costa AM, Larroche JC, Droulle P, et al. Fetal Biometry. Growth charts for practical use in fetopathology and antenatal ultrasonography. Introduction. *Fetal Diagn Ther* 1995; **10**: 211-278.

251. Hadlock FP, Deter RL, Harrist RB, et al. Fetal head circumference: relation to menstrual age. *AJR Am J Roentgenol* 1982; **138**: 649-653.

252. Hadlock FP, Deter RL, Harrist RB, et al. Fetal abdominal circumference as a predictor of menstrual age. *AJR Am J Roentgenol* 1982; **139**: 367-370.

253. Hadlock FP, Deter RL, Harrist RB, et al. Fetal biparietal diameter: a critical re-evaluation of the relation to menstrual age by means of real-time ultrasound. *J Ultrasound Med* 1982; **1**: 97-104.

254. Hadlock FP, Harrist RB, Deter RL, et al. Fetal femur length as a predictor of menstrual age: sonographically measured. *AJR Am J Roentgenol* 1982; **138**: 875-878.

255. Hoffbauer H, Pachaly J, Arabin B, et al. Control of fetal development with multiple ultrasonic body measures. *Contrib Gynecol Obstet* 1979; **6**: 147-156.

256. Issel EP. Ultrasonic measurement of the growth of fetal limb bones in normal pregnancy. *J Perinat Med* 1985; **13**: 305-313.

257. Issel EP, Prenzlau P, Bayer H, et al. The measurement of fetal growth during pregnancy by ultra-sound (B-scan). *J Perinat Med* 1975; **3**: 269-275.

258. Jeanty P, Cousaert E,Cantraine F. Normal growth of the abdominal perimeter. *Am J Perinatol* 1984; **1**: 129-135.

259. Jeanty P, Cousaert E, Cantraine F, et al. A longitudinal study of fetal limb growth. *Am J Perinatol* 1984; **1**: 136-144.

260. Jeanty P, Cousaert E, Hobbins JC, et al. A longitudinal study of fetal head biometry. *Am J Perinatol* 1984; **1**: 118-128.

261. Jeanty P, Kirkpatrick C, Dramaix-Wilmet M, et al. Ultrasonic evaluation of fetal limb growth. *Radiology* 1981; **140**: 165-168.

262. Johnsen SL, Wilsgaard T, Rasmussen S, et al. Longitudinal reference charts for growth of the fetal head, abdomen and femur. *Eur J Obstet Gynecol Reprod Biol* 2006; **127**: 172-185.

263. Jung SI, Lee YH, Moon MH, et al. Reference charts and equations of Korean fetal biometry. *Prenat Diagn* 2007; **27**: 545-551.

264. Kurmanavicius J, Wright EM, Royston P, et al. Fetal ultrasound biometry: 1. Head reference values. *Br J Obstet Gynaecol* 1999; **106**: 126-135.

265. Kurmanavicius J, Wright EM, Royston P, et al. Fetal ultrasound biometry: 2. Abdomen and femur length reference values. *British Journal of Obstetrics and Gynaecology* 1999; **106**: 136-143.

266. Lai FM,Yeo GS. Reference charts of foetal biometry in Asians. *Singapore Med J* 1995; **36**: 628-636.

267. Larsen T, Petersen S, Greisen G, et al. Normal fetal growth evaluated by longitudinal ultrasound examinations. *Early Human Development* 1990; **24**: 37-45.

268. Lei H,Wen SW. Ultrasonographic examination of intrauterine growth for multiple fetal dimensions in a Chinese population. Central-South China Fetal Growth Study Group. *Am J Obstet Gynecol* 1998; **178**: 916-921.

269. Lessoway VA, Schulzer M, Wittmann BK, et al. Ultrasound fetal biometry charts for a North American Caucasian population. *J Clin Ultrasound* 1998; **26**: 433-453.

270. Leung TN, Pang MW, Daljit SS, et al. Fetal biometry in ethnic Chinese: biparietal diameter, head circumference, abdominal circumference and femur length. *Ultrasound Obstet Gynecol* 2008; **31**: 321-327.

271. Levi S,Erbsman F. Antenatal fetal growth from the nineteenth week. Ultrasonic study of 12 head and chest dimensions. *Am J Obstet Gynecol* 1975; **121**: 262-268.

272. Levi S,Smets P. Intra-uterine fetal growth studied by ultrasonic biparietal measurements. The percentiles of biparietal distribution. *Acta Obstet Gynecol Scand* 1973; **52**: 193-198.

273. Lu SC, Chang CH, Yu CH, et al. Reappraisal of fetal abdominal circumference in an Asian population: analysis of 50,131 records. *Taiwan J Obstet Gynecol* 2008; **47**: 49-56.

274. Mathai M, Thomas S, Peedicayil A, et al. Growth pattern of the Indian fetus. *Int J Gynaecol Obstet* 1995; **48**: 21-24.

275. Merz E, Kim-Kern MS,Pehl S. Ultrasonic mensuration of fetal limb bones in the second and third trimesters.[Erratum appears in JCU 1987 Nov-Dec;15(9):704]. *Journal of Clinical Ultrasound* 1987; **15**: 175-183.

276. Munjanja SP, Masona D,Masvikeni S. Fetal biparietal diameter and head circumference measurements: results of a longitudinal study in Zimbabwe. *Int J Gynaecol Obstet* 1988; **26**: 223-228.

277. Munoz WP, Moore PJ, MacKinnon A, et al. Biparietal diameter and menstrual age in the black population attending Edendale Hospital. *J Clin Ultrasound* 1986; **14**: 681-688.

278. Neufeld LM, Haas JD, Grajeda R, et al. Ultrasound measurement of fetal size in rural Guatemala. *Int J Gynaecol Obstet* 2004; **84**: 220-228.

279. O'Brien GD,Queenan JT. Growth of the ultrasound fetal femur length during normal pregnancy. Part I. *Am J Obstet Gynecol* 1981; **141**: 833-837.

280. Okonofua FE, Ayangade SO,Ajibulu OA. Ultrasound measurement of fetal abdominal circumference and the ratio of biparietal diameter to transverse abdominal diameter in a mixed Nigerian population. *Int J Gynaecol Obstet* 1988; **27**: 1-6.

281. Okupe RF, Coker OO,Gbajumo SA. Assessment of fetal biparietal diameter during normal pregnancy by ultrasound in Nigerian women. *Br J Obstet Gynaecol* 1984; **91**: 629-632.

282. Paladini D, Rustico M, Viora E, et al. Fetal size charts for the Italian population. Normative curves of head, abdomen and long bones. *Prenat Diagn* 2005; **25**: 456-464.

283. Pang MW, Leung TN, Sahota DS, et al. Customizing fetal biometric charts. *Ultrasound Obstet Gynecol* 2003; **22**: 271-276.

284. Persson PH, Grennert L, Gennser G, et al. Normal range curves for the intrauterine growth of the biparietal diameter. *Acta Obstet Gynecol Scand Suppl* 1978; **78**: 15-20.

285. Persson PH,Weldner BM. Normal range growth curves for fetal biparietal diameter, occipito frontal diameter, mean abdominal diameters and femur length. *Acta Obstet Gynecol Scand* 1986; **65**: 759-761.

286. Pineau JC, Guihard-Costa AM,Droulle P. Only two-phase models, computed independently for males and females, are appropriate to describe fetal head growth. *Fetal Diagn Ther* 2003; **18**: 207-216.

287. Queenan JT, Kubarych SF, Cook LN, et al. Diagnostic ultrasound for detection of intrauterine growth retardation. *Am J Obstet Gynecol* 1976; **124**: 865-873.

288. Saksiriwuttho P, Ratanasiri T,Komwilaisak R. Fetal biometry charts for normal pregnant women in northeastern Thailand. *J Med Assoc Thai* 2007; **90**: 1963-1969.

289. Schluter PJ, Pritchard G,Gill MA. Ultrasonic fetal size measurements in Brisbane, Australia. *Australas Radiol* 2004; **48**: 480-486.

290. Shohat T,Romano-Zelekha O. Ultrasonographic measurements of fetal femur length and biparietal diameter in an Israeli population. *Isr Med Assoc J* 2001; **3**: 166-168.

291. Siwadune T, Sunsaneevithayakul P, Titapant V, et al. Charts of Thai fetal biometries: 2. Biparietal diameter. *Journal of the Medical Association of Thailand* 2000; **83**: 292-298.

292. Smulian JC, Ananth CV, Vintzileos AM, et al. Revisiting sonographic abdominal circumference measurements: a comparison of outer centiles with established nomograms. *Ultrasound Obstet Gynecol* 2001; **18**: 237-243.

293. Snijders RJ,Nicolaides KH. Fetal biometry at 14-40 weeks' gestation. *Ultrasound Obstet Gynecol* 1994; **4**: 34-48.

294. Sunsaneevithayakul P, Boriboonhirunsarn D, Siwadune T, et al. Charts of Thai fetal biometries: 4. Abdominal circumference. *Journal of the Medical Association of Thailand* 2000; **83**: 307-314.

295. Tamura RK,Sabbagha RE. Percentile ranks of sonar fetal abdominal circumference measurements. *Am J Obstet Gynecol* 1980; **138**: 475-479.

296. Thame M, Osmond C, Fletcher H, et al. Ultrasound derived fetal growth curves for a Jamaican population. *West Indian Med J* 2003; **52**: 99-110.

297. Titapant V, Siwadune T, Boriboonhirunsarn D, et al. Charts of Thai fetal biometries: 3. Femur length. *Journal of the Medical Association of Thailand* 2000; **83**: 299-306.

298. Todros T, Ferrazzi E, Groli C, et al. Fitting growth curves to head and abdomen measurements of the fetus: a multicentric study. *J Clin Ultrasound* 1987; **15**: 95-105.

299. Verburg BO, Steegers EA, De Ridder M, et al. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. *Ultrasound Obstet Gynecol* 2008; **31**: 388-396.

300. Warda AH, Deter RL, Rossavik IK, et al. Fetal femur length: a critical reevaluation of the relationship to menstrual age. *Obstet Gynecol* 1985; **66**: 69-75.

301. Watson DS. Biparietal diameter in the Australian aboriginal fetus. *Br J Obstet Gynaecol* 1986; **93**: 339-342.

302. Westerway SC, Davison A,Cowell S. Ultrasonic fetal measurements: new Australian standards for the new millennium. *Aust N Z J Obstet Gynaecol* 2000; **40**: 297-302.

303. Wladimiroff JW, Bloemsma CA,Wallenburg HC. Ultrasonic assessment of fetal head and body sizes in relation to normal and retarded fetal growth. *Am J Obstet Gynecol* 1978; **131**: 857-860.

304. Gardosi J. Customized growth curves. *Clin Obstet Gynecol* 1997; **40**: 715-722.

305. Gardosi J. Dating of pregnancy: time to forget the last menstrual period. *Ultrasound Obstet Gynecol* 1997; **9**: 367-368.

306. Mongelli M, Wilcox M,Gardosi J. Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. *Am J Obstet Gynecol* 1996; **174**: 278-281.

307. Salomon LJ. Early fetal growth: concepts and pitfalls. *Ultrasound Obstet Gynecol* 2010; **35**: 385-389.

308. Sarris I, Ioannou C, Dighe M, et al. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. *Ultrasound in Obstetrics & Gynecology* 2011; **38**: 681-687.