THE STATISTICAL AND ANALYTICAL BASIS OF DOWN SYNDROME SCREENING

AND

AN EVALUATION OF DOWN SYNDROME SCREENING IN GWENT AND SOUTH GLAMORGAN

A THESIS SUBMITTED TO
THE UNIVERSITY OF LEEDS MEDICAL SCHOOL
IN COMPLETION OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF MEDICINE

BY

TIMOTHY MARK REYNOLDS
BSc (Leeds) MB ChB (Leeds) MRCPath

AUGUST 1993
<table>
<thead>
<tr>
<th></th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>ii</td>
<td>Declaration</td>
</tr>
<tr>
<td>iii</td>
<td>Acknowledgements</td>
</tr>
<tr>
<td>iv</td>
<td>Index</td>
</tr>
<tr>
<td>xi</td>
<td>Abbreviations</td>
</tr>
<tr>
<td>xii</td>
<td>List of Figures</td>
</tr>
<tr>
<td>xiv</td>
<td>List of Tables</td>
</tr>
<tr>
<td>xvii</td>
<td>List of Mathematical Formulae</td>
</tr>
<tr>
<td>xviii</td>
<td>List of Appendices</td>
</tr>
<tr>
<td>xix</td>
<td>Summary</td>
</tr>
</tbody>
</table>
Declaration

I, Timothy Mark Reynolds, declare that this thesis is the result of my independent investigations, unless otherwise stated. I also certify that the work referred to in this thesis has not been accepted for any degree or submitted elsewhere in candidature for any degree.

........................................
Timothy Mark Reynolds
ACKNOWLEDGEMENTS

I would like to acknowledge the following people who have assisted and encouraged me during the preparation of this thesis.

Mr. C. Davies, Senior Project Leader, Kodak Diagnostics Ltd.
Mr. A. Dawson, Senior Lecture in Obstetrics, UHW.
Miss C. Donovan, PhD student, Plymouth University.
Dr. F. Dunstan, Lecturer in Mathematics, Cardiff University.
Prof. G. Elder, UHW.
Mr. R Henley, Senior MLSO, UHW.
Mr. H. Hughes, Biochemist, Royal Gwent Hospital.
Dr. R. John, Biochemist, UHW.
Ms. G. Jones, midwife, UHW.
Dr. J. Macri, NTD laboratories, New York.
Dr. M. Matharu, Consultant in Public Health, Gwent H. A.
Dr. B. Nix, Lecturer in Mathematics, Cardiff University.
Mr. D. O’Sullivan, computer manager, Royal Gwent Hospital.
Dr. B. Passass, Abbott Diagnostics Ltd.
Dr. M. Penney, Consultant Chemical Pathologist, Royal Gwent Hospital.
Mr. K. Spencer, Oldchurch Hospital, Romford.
Dr. P. Wood, Southampton General Hospital.
Dr. D. Wright, Principal Lecturer in Mathematics, Plymouth University.
The MLSO staff of the Royal Gwent Hospital and UHW.

Finally I would also like to thank Abbott Diagnostics Ltd. for provision of the computer which allowed me to continue my research, and my wife for her understanding when I spent my evenings writing.
INDEX

chapter 1:

INTRODUCTION TO DOWN SYNDROME 1

1.1 HISTORICAL ASPECTS 2

  1.1.1 Down Syndrome in Antiquity 2
  1.1.2 Early Descriptions of Down Syndrome 3
  1.1.3 Early Ideas on Causation (Non-Genetic) 5
  1.1.4 Development of a Genetic Theory of Causation 8

1.2 GENETICS OF DOWN SYNDROME 11

  1.2.1 Non-Familial Down Syndrome 11
  1.2.2 Familial Down Syndrome 14

1.3 CHARACTERISTICS OF DOWN SYNDROME INDIVIDUALS 17

  1.3.1 Major Features 17
  1.3.2 Occasional Features 19
  1.3.3 Diagnostic Features in the Neonate 19
  1.3.4 Natural History of Down Syndrome 20

1.4 EPIDEMIOLOGY OF DOWN SYNDROME 20

  1.4.1 Incidence of Down Syndrome 20
  1.4.2 Survival of Down Syndrome Individuals 24

1.5 HEALTH ECONOMICS OF DOWN SYNDROME 26

chapter 2:

INTRODUCTION TO DOWN SCREENING 32

2.1 GENERAL PRINCIPLES OF SCREENING 33

  2.1.1 Definitions 33
  2.1.2 Ethical Considerations 36
2.2 Development of Down Syndrome Screening 40

2.3 The Meaning of Risk 49

Chapter 3:

General Methods 55

3.1 Biochemical Assay Methods 57

3.1.1 Royal Gwent Assays 57
3.1.1.1 AFP 57
3.1.1.1.1 AFP-free serum preparation 59
3.1.1.2 total HCG 60
3.1.2 UHW Assays 62
3.1.2.1 Magnetic Solid Phase Preparation 62
3.1.2.2 AFP 63
3.1.2.3 total HCG 64
3.1.3 Amersham Assays 64
3.1.3.1 AFP 65
3.1.3.2 total HCG 66
3.1.3.3 uE3 2nd trimester assay 67
3.1.3.4 uE3 3rd trimester assay 68
3.1.3.5 uE3 3rd trimester assay, Canick / Wald's modification 69
3.1.4 CIS free β HCG assay 69

3.2 Mathematical Methods 70

3.2.1 Age Related Risk Calculation 70
3.2.2 Calculation of Multiples of the Median (MoM) 71
3.2.3 Weight Correction of MoMs 74
3.2.4 Risk Calculation Methods 75
3.2.4.1 Linear algorithm 75
3.2.4.2 Discriminant Function Method 75
3.2.4.3 Likelihood Ratio Method 76
3.2.5 Statistical and Matrix Functions 78
3.2.5.1 Gaussian Distribution Formulae 78
3.2.5.2 Matrix Mathematics 79
3.2.5.2.1 Matrix Notation and Orientation 80
3.2.5.2.2 Matrix Addition and Subtraction 80
3.2.5.2.3 Matrix Multiplication 81
3.2.5.2.4 Determinants 82
3.2.5.2.5 Matrix Inversion 84
3.3 Simulation Methods

3.3.1 Random Normal Deviate Method
3.3.2 Cholesky Root Method for Correlated Variables

3.4 Statistical Methods

3.4.1 Deming’s Regression
3.4.2 Skewness and Kurtosis
3.4.3 Confidence Intervals for Medians
3.4.4 Kolgomorov-Smirnov-type tests

3.5 Ultrasound Formulae

Chapter 4:

Population Norms and Medians

4.1 Population Normal Ranges

4.1.1 Source of data
4.1.2 Age distribution
4.1.3 Weight distribution
4.1.4 Height distribution
4.1.5 Biparietal diameter distribution
4.1.6 AFP distribution
4.1.7 HCG distribution
4.1.8 Distributions of other analytes

4.2 Derivation of Medians

4.2.1 Methods for derivation of medians
4.2.2 Weekly medians
4.2.3 Weekly weighted regressed medians
4.2.4 BMUS dated, regressed medians
4.2.5 Comparison of different medians

4.3 Distribution of Multiples of the Median

4.3.1 MoM distributions for different medians
4.3.2 Effectiveness of correction for gestation age
4.3.3 Normality of MoM distributions
4.3.4 Correlation of AFP MoMs versus HCG MoMs
chapter 5:

MATERNAL FACTORS AFFECTING MULTIPLES OF THE MEDIAN AND DOWN SYNDROME RISKS

5.1 MATERNAL WEIGHT

5.1.1 The Rationale for Weight Correction
5.1.2 Preliminary Assessment of the Need for Weight Correction
5.1.3 The Effect of Weight correction in a Large Sample

5.2 MATERNAL HEIGHT

5.3 MATERNAL DIABETES

5.4 MATERNAL RACE

5.5 CONCLUSIONS

chapter 6:

THE IMPORTANCE OF ACCURATE GESTATION DATING IN DOWN SYNDROME SCREENING

6.1 THE PROBLEM OF GESTATION DATING

6.2 DEMONSTRATION OF ULTRASOUND IMPRECISION

6.3 DEMONSTRATION OF THE EFFECT OF ULTRASOUND IMPRECISION ON SCREENING

6.3.1 Demonstration of Edge Effects due to Ultrasonic Dating
6.3.1.1 The Significance of Edge Effects
6.3.2 Demonstration of ’Quantum’ Effects due to
chapter 7:

LABORATORY INFLUENCES IN DOWN SYNDROME SCREENING 185

7.1 THE IMPORTANCE OF ASSAY OPTIMISATION 186

7.1.1 Comparison of a poorly optimised assay (IM2-uE3 Canick/Wald modification) with a fully optimised assay (IM4-uE3). 186
7.1.1.1 Comparison of Assay Standardisation 187
7.1.1.2 Comparison of Assay Results 190
7.1.1.3 The Effect of Conversion to MoMs 191
7.1.1.4 The Effect on Likelihood Ratios 193
7.1.1.5 The Effect on Assay Imprecision 197
7.1.1.6 Conclusions 197

7.1.2 Comparison of two optimised AFP assays 201

7.2 THE EFFECT OF ASSAY IMPRECISION 203

7.3 COMBINATION OF ASSAY AND DATING IMPRECISION 208

7.4 CONCLUSIONS 212

chapter 8:

STATISTICAL ASPECTS OF DOWN SCREENING 214

8.1 THE EFFECT OF MATERNAL AGE ON SCREENING 215

8.1.1 Derivation of age effects 215
8.1.2 Conclusions 224

8.2 POPULATION PARAMETERS 225

8.2.1 What values should we use? 225
8.2.2 Simulation-based evaluation of the most suitable parameters 226
8.2.3 Data-based evaluation of the most suitable parameters 235
8.2.4 Conclusions

8.3 Which Analytes Are Useful?

8.3.1 The second trimester
8.3.1.1 Is Free $\beta$ HCG Useful?

8.3.2 The first trimester

8.4 Other Risk Calculation Algorithms

8.4.1 Assessment of atypicality
8.4.2 An examination of the discriminant function method of Nørgaard-Pedersen et al, (1990)
8.4.3 An examination of the calculation method of Crossley et al, (1991)
8.4.4 An examination of the method of Muller and Boué, 1990

8.5 The Effect Of Weight Correction On Detection

8.6 Conclusions

Chapter 9:

Evaluation of Down Syndrome Screening in Gwent and South Glamorgan

9.1 Prediction Of Down Syndrome Incidence

9.2 One Year's Experience Of Down Syndrome Screening

9.2.1 Description of Study
9.2.2 Results of Study
9.2.3 Discussion
9.2.4 Inter-Disciplinary Concerns About Screening

Chapter 10.

Conclusions And Prospects For The Future
10.1 DISCUSSION

10.1.1 Why should we screen for Down syndrome? 289
10.1.2 What method of screening should we use? 290
10.1.3 Which analytes should we use? 291
10.1.4 How should we standardise for gestation age? 292
10.1.5 What other maternal factors should be taken into account? 293
10.1.6 What is the significance of gestation dating? 295
10.1.7 Which mathematical algorithm should we use? 295
10.1.8 What can we expect from Down syndrome screening? 295
10.1.9 What are the concerns about Down syndrome screening? 296

10.2 FUTURE AVENUES IN SERUM SCREENING 298

10.2.1 Correlation between analytes may not be bad 298
10.2.2 Correlation between analytes may be spurious 302
10.2.3 Gestation dating may affect medians 302
10.2.4 Risk calculation algorithms 303
10.2.5 Ethnic differences 303
10.2.6 Inter-kit variations in population parameters 303
10.2.7 Within-kit variability 304
10.2.8 Within patient variation 304

10.3 FUTURE AVENUES IN DOWN SCREENING 305

10.3.1 Fetal cells in the maternal circulation 305
10.3.2 FISHing for a diagnosis 307
10.3.3 PCR possibilities 308

BIBLIOGRAPHY 309

APPENDICES 360
ABBREVIATIONS

AFP  $\alpha$-fetoprotein
BPD  Biparietal diameter
CRL  Crown-rump length
FACS Fluorescence activated cell sorting
FISH Fluorescent in-situ hybridisation
FL   Femur length
FN   False negative
FP   False positive
HCG  Human choriogonadotropin
IM4-AFP Amerlex-M second trimester AFP assay
IM4-HCG Amerlex-M second trimester HCG assay
IM4-uE3 Amerlex-M second trimester uE3 assay
IM2-uE3 Modified Amerlex-M third trimester uE3 assay
IRMA Immunoradiometric assay
MoM Multiple of the median
NTD Neural tube defect
PCR Polymerase chain reaction
RGH Royal Gwent Hospital
RGH-AFP Royal Gwent Hospital AFP assay
RGH-HCG Royal Gwent Hospital HCG assay
RIA Radioimmunoassay
ROC Receiver-operator curve
TN  True negative
TP  True positive
uE3  Unconjugated oestriol
UHW University Hospital of Wales
UHW-AFP University Hospital of Wales AFP assay
UHW-HCG University Hospital of Wales HCG assay
**List of Figures**

1.1 Effect of non-disjunction in meiosis.
1.2 Possible gametes from a translocation carrier.
1.3 Survival of Down syndrome children up to age 10.
2.1 Risk and dread.
3.1 Change in AFP, HCG and uE3 concentrations with gestation age.
3.2 Plot of AFP MoM distributions (Palomaki and Haddow, 1987).
4.1 Request for for Down syndrome screening.
4.2 Age distribution of pregnant population of Gwent.
4.3 Weight distribution of pregnant population of Gwent.
4.4 Height distribution of pregnant population of Gwent.
4.5 Relationship between weight and height.
4.6 Fetal BPD distribution.
4.7 Graph to show different medians for AFP.
4.8 Graph to show different medians for HCG.
4.9 Probability plot for AFP MoMs.
4.10 Probability plot for HCG MoMs.
6.1 Method for derivation of $P_{\text{incorrect\ week}}$ for ultrasound dating.
6.2 Probability of incorrect gestation week related to ultrasound imprecision.
6.3 Comparison of likelihood ratios derived from weekly and BPD based MoMs (BPD base = 34 mm).
6.4 Comparison of likelihood ratios derived from weekly and BPD based MoMs (BPD base = 36 mm).
7.1 Dose response curves for IM2 amd IM4 uE3 standards assayed with IM4 assay.
7.2 Dose response curves for IM2 amd IM4 uE3 standards assayed with IM4 assay.
7.3 Comparison of uE3 patient results assayed with IM2 and IM4 kits.
7.4 Comparison of uE3 MoMs assayed with IM2 and IM4 kits.
7.5 Comparison of likelihood ratios calculated from IM2 and IM4 MoMs.
7.6 Comparison of precision of IM2 and IM4 kits.
7.7 Effect of assay and dating imprecision on likelihood ratios (integer week dating).
7.8 Effect of assay and dating imprecision on likelihood ratios (BPD derived dating).
8.1 ROC plot of detection and false positive rate for different screening policies.
8.2 ROC plot of data from section 8.2.2 from Wald et al, 1988a.
8.3 ROC plot of data from section 8.2.2 from Wald et al, 1992a.
8.4 ROC plot of 'double' and 'triple' testing using 'real' data.
8.5 Diagram to show why 'triple' testing causes loss of detection.
8.6 Interpretation algorithm for assessment of atypicality.
8.7 Demonstration of Atypicality.
8.8 ROC plot comparing likelihood ratio method with discriminant function method.
8.10 ROC plot comparing discriminant function method with method of Crossley et al.
8.11 ROC plot comparing likelihood ratio method with method of Muller and Boué, 1990.
8.12 ROC plot comparing methods of Crossley et al and Muller and Boué.
8.13 ROC plot showing effect of weight correction.
9.1 Summary of one year's experience in South Wales.
10.1 The effect of analyte correlation, r = 0.0
10.2 The effect of analyte correlation, r = 0.5
10.3 The effect of analyte correlation, r = 0.9
LIST OF TABLES

1.1 Examples of chromosomal abnormalities due to non-disjunction.
1.2 Retrospective studies of Down syndrome incidence.
1.3 Prospective studies of Down syndrome incidence.
1.4 Birth incidences of some other congenital conditions.
1.5 Life table for Down syndrome individuals.
1.6 Costs and benefits of Down syndrome screening.
1.7 Effectiveness of AFP screening for chromosomal abnormalities.
1.8 Effectiveness of AFP/HCG screening.
3.1 Age related risk, regression coefficients.
4.1 Age distribution of pregnant population of Gwent.
4.2 Weight distribution of pregnant population of Gwent.
4.3 Height distribution of pregnant population of Gwent.
4.4 Distribution of BPD's and relationship to gestation age.
4.5 AFP concentration distribution.
4.6 HCG concentration distribution.
4.7 Percentile distribution of AFP and HCG at different gestation ages.
4.8 Statistical details of distributions of AFP HCG.
4.9 Weighted least squares regression parameters for gestation week vs AFP and HCG concentration.
4.10 Least squares regression parameters for BMUS formula derived gestation age vs AFP and HCG concentration.
4.11 Medians for weeks 15 - 19 derived by 5 methods.
4.12 AFP MoM distributions using the medians in table 4.11.
4.13 HCG MoM distributions using the medians in table 4.11.
5.1 Median serum AFP, HCG and uE3 MoMs related to maternal weight.
5.2 Regression parameters for MoMs vs weight.
5.3 Effect of weight correction on ‘screen positive’ rate.
5.4 Median serum AFP, HCG and uE3 MoMs related to maternal weight in data set 1.
5.5 Regression parameters for MoMs vs weight in dataset 1.
5.6 Distribution of uncorrected and weight corrected MoMs for AFP and HCG in dataset 1.
5.7 Comparison of Gwent population parameters with published parameters.
5.8 ‘Screen positive’ rates for uncorrected and weight corrected data for dataset 1.
5.9 Correlation between weight, height and AFP MoMs.
5.10 Regression coefficients for weight / height correction.
5.11 Correlation between weight and height corrected MoMs and weight and height.
5.12 ‘Screen positive’ rates for weight and height corrected data.
6.1 Accuracy of determination of gestation age.
6.2 BPD size bands corresponding to gestation week.
6.3 Exponential regression coefficients for BPD vs AFP and HCG.
6.4 Medians for extremes of BPD size bands.
6.5 Comparison of BPD based with week based MoMs.
6.6 Variation in Down syndrome risks at the middle and extremes of size bands.
6.7 Comparison of week based with BPD based MoMs.
6.8 Variation in Down syndrome risks with gestation week.
6.9 Distribution of likelihood ratios for different weekly medians (BPD base = 34 mm).
6.10 Distribution of likelihood ratios for different weekly medians (BPD base = 36 mm).
6.11 Distribution of likelihood ratios for different BPD based medians (BPD base = 34 mm).
6.12 Distribution of likelihood ratios for different BPD based medians (BPD base = 36 mm).
7.1 Distribution of MoMs for uE3 results from IM2 and IM4 kits.
7.2 Distribution of MoMs for AFP results from Amersham and UHW AFP assays.
8.1 Age related risk of Down syndrome and likelihood ratio needed to modify to risk cut off.
8.2 Age related detection rates.
8.3 Age related false positive rated.
8.4 Age related predictive value of a positive rate.
8.5 Population parameters for unaffected pregnancies.
8.6 Population parameters for Down syndrome pregnancies.
8.7 Proportions of simulated Down syndrome population with the same likelihood ratio as unaffected population for a 'double' test.
8.8 Proportions of simulated Down syndrome population with the same likelihood ratio as unaffected population for a 'triple' test.
8.9 Detection and false positive rates for 'double' and 'triple' testing.
8.10 Comparison of screening with total HCG and free β HCG.
8.11 Down syndrome risks and Mahalanobis distances for 37 non-Down syndrome pregnancies.
8.12 Summary of effects of using atypicality assessment.
8.13 Detection and false positive rates for screening using discriminant function method.
8.16 Detection and false positive rates for screening using method of Muller and Boué, 1990.
8.17 Detection and false positive rates for screening with and without weight correction.
9.1 Prediction of expected number of Down syndrome cases per year in the populations served by The Royal Gwent and UHW
screening programs.

9.2 Down syndrome cases identified during the prospective trial.
9.3 Amniocenteses performed during the prospective trial.
**List of Mathematical Formulae**

3.1 Age related risk formula
3.2 Down screening algorithm (DiMaio *et al*, 1987).
3.5 Modification of Age risk by likelihood ratio.
3.6 Univariate Gaussian distribution function.
3.7 Bivariate Gaussian distribution function.
3.8 Trivariate Gaussian distribution function.
3.9 Matrix addition and subtraction.
3.10 Matrix multiplication.
3.11 Definition of determinant of a matrix.
3.12 Special cases for determinant of (2 x 2) and (3 x 3) matrices.
3.13 Expansion by cofactors for inverse matrix determination.
3.14 Inversion of a (2 x 2) matrix.
3.15 Inversion of a (3 x 3) matrix.
3.16 Covariance matrix (V), its cofactor matrix and determinant.
3.18 Calculation of p given standard normal deviate.
3.19 Comparison between least squares and Deming’s regressions.
3.20 Calculation of coefficients of skewness and kurtosis.
3.21 Calculation of confidence limits for medians.
3.22 Critical value cut-offs for Smirnov and Lilliefors tests.
3.23 Ultrasound formulae.
8.1 Baye’s theorem.
8.2 Calculation of Mahalanobis distances.
8.3 Double analyte discriminant function method (Nørgaard-Pedersen *et al*, 1990).
LIST OF APPENDICES

1. PASCAL subroutines.
2. List of suppliers.
Screening for Down syndrome in early pregnancy by combining the results of multiple serum markers in a mathematical algorithm was first described in 1988 and expanded on the technique for screening using maternal age and AFP described in 1987. The multiple marker technique purported to be able to offer a significant increase in detection whilst maintaining the amniocentesis rate unchanged. The experience in South Wales with a 'double' test of AFP and total HCG confirms this claim as there was a significant increase in the number of cases of Down syndrome that were identified, compared to the historical detection rate, without any significant increase in amniocentesis requests.

This thesis examines the statistical basis behind the multiple marker test and demonstrates that: For a laboratory starting a new screening program, the likelihood ratio method is the best algorithm for risk calculation; Gestation dating must be carried out to the nearest day and this should preferably be done using ultrasound confirmed dates; Medians for derivation of multiples of the median must be derived either by weighted regression of log transformed week-derived medians or by least squares regression of all data (after log transformation); Correction for maternal weight improves detection; All assays must be completely optimised for the analyte concentrations expected (failure to do so may result in biases that may seriously diminish the effectiveness of screening); Assay imprecision should be kept as small as possible as this affects the number of assays that can be used to derive risk estimates; the current estimates of screening efficiency, based on mathematical simulations are unsatisfactory because simulations can lead to estimates
wildy at variance with reality (large prospective studies are needed to confirm the benefit of Down screening); Despite the effectiveness of the screen there are still some areas of concern, particularly psychological effects on those young women who had never previously considered Down syndrome and who receive false positive results.

Overall it is concluded that serum screening for Down syndrome is worthwhile but that in the long term other technologies may supersede. For example, extraction of pure samples of fetal cells from maternal blood that may then be tested for Down syndrome (and other trisomies) by fluorescent in-situ hybridisation (FISH) techniques or by quantitative polymerase chain reaction (PCR) techniques. All of these methods have been tested in research settings and have been found to be viable: all that is required is a method of rendering them suitable for use in a routine laboratory.
chapter 1:

INTRODUCTION TO DOWN SYNDROME

1.1 HISTORICAL ASPECTS
   1.1.1 Down Syndrome in Antiquity
   1.1.2 Early Descriptions of Down Syndrome
   1.1.3 Early Ideas on Causation (Non-Genetic)
   1.1.4 Development of a Genetic Theory of Causation

1.2 GENETICS OF DOWN SYNDROME
   1.2.1 Non-Familial Down Syndrome
   1.2.2 Familial Down Syndrome

1.3 CHARACTERISTICS OF DOWN SYNDROME INDIVIDUALS
   1.3.1 Major Features
   1.3.2 Occasional Features
   1.3.3 Diagnostic Features in the Neonate
   1.3.4 Natural History of Down Syndrome

1.4 EPIDEMIOLOGY OF DOWN SYNDROME
   1.4.1 Incidence of Down Syndrome
   1.4.2 Survival of Down Syndrome Individuals

1.5 HEALTH ECONOMICS OF DOWN SYNDROME
1.1 Historical Aspects
(Benda, 1969: Smith and Berg, 1976: Brousseau, 1928)

1.1.1 Down Syndrome in Antiquity

Prior to descriptions in the medical literature, there are a number of paintings in which the subjects appear to show the stigmata of Down syndrome: ‘Madonna and Child’ (15th century) by Andrea Mantegna (Rührah, 1935) in which the child shows some features of cretinism (coarse hair, prominent tongue, thickened mucous membranes and oedematous appearance) and some of Down syndrome (a curved little finger, slanted eyes, a small nose, an open mouth and adenoidal expression); ‘The Peasant and the Satyr’ (1635-40) by Jacob Jordaens (Zellweger, 1968); ‘Woman with child’ (1616) by Jacob Jordaens (Zellweger, 1968): ‘Lady Cockburn and her Children’ (1773) by Joshua Reynolds (Zellweger, 1968). Interpretation of such visual evidence is however fraught with difficulty as the child in the Joshua Reynolds painting is Sir George Cockburn who later became Admiral of the British fleet and commanded the HMS Northumberland when she carried Napoleon Bonaparte to his final exile on St Helena in 1815 (Zellweger, 1968).

In addition to the artistic representations of Down syndrome there is archaeological evidence from the late Saxon burial ground at Breedon - on - the - Hill in Leicestershire. Brothwell (1960) described the skull of a child aged approximately 9 years which he believed was from a child with Down syndrome. The rarity of ancient Down syndrome skeletons is unsurprising because of the smaller populations with a younger age structure and higher infant mortality than in modern times. For example,
it has been estimated that when Shakespeare was born (1584) there may have been only 100 Down syndrome individuals in the whole of Great Britain (Richards, 1968).

1.1.2 Early Descriptions of Down Syndrome

The condition now known as Down (or Down’s) syndrome was first described as a definite entity in 1866 when J. Langdon Down published his classic paper “Ethnic Classification of Idiots” in which he described the mongoloid idiot as:

. . . a representative of the great mongolian race: When placed side by side it is difficult to believe that the specimens compared are not children of the same parents. The hair is not black, as in the real mongol, but of a brownish colour, straight and scanty. The face is flat and broad and destitute of prominence. The cheeks are roundish and extended laterally. The eyes are obliquely placed and the internal canthi more than normally distant from one another. The palpebral fissure is very narrow. The forehead is wrinkled transversely from the constant assistance which the levatores palpebrarum derive from the occipito-frontalis muscles in the opening of the eyes. The lips are large and thick with transverse fissures. The tongue is long, thick and much roughened. The nose is small. The skin has a slight dirty yellowish tinge and is deficient in elasticity, giving the appearance of being too large for the body (Langdon Down, 1866).

Also in 1866, Eduoard Séguin described mongolism, but he believed it was a sub-group of cretinism:

. . . The lowland cretinism of Belgium, Virginia - with its distinct goitre, its gray and dirty straw-coloured skin, bears the same relation to idiocy and imbecility as the more extensive alpine variety. So does ‘Furfuraceous
Cretinism’ with its milk-white rosy and peeling skin; with its shortcomings of the integuments, which gives an unfinished aspect to the truncated fingers and nose; with its cracked lips and tongue; with its red, ectopic conjunctiva, coming out to supply the curtailed skin at the margin of the lids (Séguin, 1866).

The first scientific report on Down syndrome was made at a meeting of the Medico-Psychological Association in 1875 and was published in 1876. This report named the syndrome ‘Kalmuc Idiocy’ (after the Kalmuc region of Russia) and noted a lack of hereditary factors, no kinship between parents of affected infants, a tendency for increased maternal age and brachycephaly but did not mention Down’s paper of 1866 (Fraser and Mitchell, 1876). Later, several authors credited Down with the discovery of the syndrome (Ireland, 1877; Shuttleworth, 1883; Kovalevsky, 1906) because although there were earlier descriptions (Séguin, 1846, 1866; Esquirol, 1838), these had failed to identify it as a unique entity.

By 1877, ‘Mongolism’ had become a recognised type of mental deficiency (Ireland, 1877) and Down syndrome individuals were described as ‘unfinished children’ meaning that some factor had depressed the maternal powers so that development was incomplete (Shuttleworth, 1883, 1886).

... My view is that they are in fact unfinished children and that their peculiar appearance is really that of a phase of foetal life. I do not mean that they are necessarily prematurely born, but that some cause has depressed the maternal powers, and there has been a defect of formative force. It is remarkable that, in my experience, nearly one-half of these children are the last born of a long family and in more than one-third a phthsical [tuberculous] history has been traced (Shuttleworth, 1886).
Many characteristics of mongoloid children were described during this period: the mouth and jaws (Jones R, 1890); the eyes (Oliver 1891); malposition; irregularities of contour and inequality in comparative size of the bony orbits, with obliquity of the attached ligamental and tarsal tissues, giving the palpebral fissures their peculiarity of direction, the lids their shortness. . . . the ocular bulb in nearly every instance presents peculiarities of structural change characterised by the appearances of the results of low and chronic forms of neuro-retinitis and choroiditis indicative of local inflammation of those parts before and after birth of the individual (Oliver, 1891).

the hands, notably curvature of the little finger (Smith T, 1896); the association with congenital heart disease (Garrod, 1894, 1898); and the brain (Wilmarth, 1890). Later, surveys covering other aspects of the condition were reported: the survival rates for Down syndrome children (Rosenburg, 1924); familial details including ABO blood groups (Orel, 1927); maternal ages and familial recurrence rates (Turpin and Caratzali, 1934: Lahdensuu, 1937: Doxiades and Portius, 1938).

1.1.3 Early Ideas on Causation (Non-genetic)

Down perceived that ‘mongolism’ had a biological cause which required a special explanation. He believed in Darwin’s theory of evolution and thus if a disease could break down racial barriers it helped to demonstrate the unity of the human species. His theory of ‘ethnic regression’ never became popular but despite this the terms ‘Mongolian’ and ‘Mongol’ came into general use. Most authorities admitted that there was no true resemblance to Mongolian individuals (Tredgold, 1908) but some persisted in the concept of reversion to earlier phylogenetic type,
especially Crookshank (1924) who thought that Down syndrome was regression not merely to a primitive oriental human type but to the orang utan. This belief was based on the presence of the palmar ‘simian’ flexion crease of the hands in patients with the condition.

. . . Amongst the many points to which I would call attention are the shape of the skull, the primitive state of air cells, the slight brow ridges and the facial bones; the eyes, the ears, the lips, tongue, hair, skin and mucous membranes, the larynx, the heart, intestines and external genitals, the hands, feet and vertebral column. The vocal peculiarities, the habitual adoption of the Buddha position, the auditory rather than visual mind and brain, the relative proportions of the cerebrum, cerebellum, pons and medulla, are all interesting and important. . . . If this is a case of reversion, it must be a reversion to a type even further back than the Mongol stock from which some ethnologists believe all the various races of men have sprung (Crookshank, 1924).

Shuttleworth (1906) and Down had also observed that familial tuberculosis was commoner in families with a ‘mongol’ child and this was considered a possible cause particularly as patients with Down syndrome frequently died of miliary tuberculosis (Tredgold, 1908: Potts, 1909). Shuttleworth (1909) made an aetiological study of 350 cases and demonstrated that there was an association between advanced age of the mother at the time of gestation and that Down individuals were often members of large families.

. . . It would seem to be a fair inference from the figures just cited that more than half of the Mongolian imbeciles in institutions are the last-born children mostly of long families and that in a considerable proportion, from one-half to one-third, the mothers were at the time of gestation
approaching the climacteric period and that in consequence, the reproductive powers were at low ebb. Which of the factors - the advanced age of the mother or her exhaustion by a long series of previous pregnancies, is the more potent factor is left in doubt. Both act concurrently in most cases (Shuttleworth, 1909).

He was unable to decide whether the condition was due to advanced maternal age or reproductive exhaustion due to a long series of pregnancies but rejected syphilis as a cause, which had been suggested by other authors (Sutherland, 1899: Hjorth, 1907). Other theories of causation included parental alcoholism (Cafferata, 1909); epilepsy, insanity, nervous instability and mental retardation in close relatives (Tredgold, 1908: Caldecott, 1909); maternal emotional distress during pregnancy; thyroid deficiency (Stoeltzner, 1919); foetal hyperthyroidism (Clark, 1929, 1933);

. . . The action hyperthyroidism has on frog embryos, including the cutting out of normal development and growth, and the action thyroxin has in retarding cell division and embryonic development would explain the general arrest of growth and development of the mongol, including that of the skull and brain (Clark, 1929).

hypoplastic foetal adrenal glands (Vas, 1925); dysfunctional foetal pituitary gland (Benda and Bixby, 1939); foetal thymic abnormality (Barnes, 1923); contraction of the amniotic sac during week 6 - 7 preventing normal development and causing dwarfism (Jansen, 1921).

. . . The mongoloid idiot is undersized. There is micromelia with shortness of hands and feet. The fingers and toes are too short. The line connecting the tip of the toes is almost at right angles to the axis of the front. In short
the extremities of the Mongoloid idiot show some or all of the characteristics of the Achondroplast. Thus an intimate relationship appears to exist between Mongolian idiots and Achondroplasia. And this is also patent from the fact that Mongoloid idiots and Achondroplastics show an affinity for the same congenital malformations, as harelip, micrognathia, clubfoot and others (Jansen, 1921).

It had been realised by Down that mongolism started very early in life and this view was reinforced by the association between Down syndrome and heart defects (Garrod, 1894, 1898, 1899: Thomson, 1898). The discovery that Down syndrome patients had characteristic dermatoglyphic patterns pin-pointed the time of maldevelopment to before the 10th week of gestation as dermal ridge patterns are fixed by this time (Cummins, 1936, 1939). This was interpreted as meaning that a genetic cause for the syndrome was likely.

1.1.4 Development of a genetic theory of causation.

A genetic cause had been postulated by some workers who believed that the 'regressive features' of 'Mongolism' were simply the result of recessive genetic characteristics concentrating in the affected individual (Herrmann, 1925). Recognition of Down syndrome in other races was a further early pointer to a genetic cause. Cases were noted in Negro’s (Bleyer, 1925: Brahdy, 1927: Dunlap, 1933), American Indians (Bleyer, 1932, 1934), Chinese and Japanese (Tumpeer, 1922: Sweet, 1934), Indians and Malaysians (Chand, 1934: Illing, 1939) and later, the Bantu in South Africa (Lötter, 1955: Kaplan, 1955).

The relationship between parental age and the incidence of Down
syndrome was independently investigated by Penrose (1933) and Jenkins (1933). It was demonstrated that paternal age was insignificant and after exclusion of the effect of maternal age on birth order (Penrose, 1934), the maternal age itself was shown to be the most important predictor of Down syndrome incidence. The first accurate estimations of incidence of Down syndrome were provided by Jenkins (1933) and Malpas (1937) at approximately 1 case per 700 live births. This is relatively common and thus the significance of a family with more than one affected child was difficult to assess. There were however, three facts which confirmed the genetic origin of the syndrome and that there were 2 distinct mechanisms of causation: Firstly, it had been observed that monozygotic twins were always equally affected whereas dizygotic twins were unequally affected (Halbertsma, 1923; Reuben and Klein, 1926). Also, a few examples of definitely dizygotic twins were found that were both affected by Down syndrome (Keay, 1958). Secondly, in some cases, the syndrome could be transmitted directly from mother to child (Lelong et al, 1949; Rehn and Thomas, 1957) and thirdly, when more than one member was affected, the dependence on maternal age was diminished (Penrose, 1951, 1953).

By studying normal testicular tissue Painter (1921) estimated the normal chromosome number as 45 - 48 and noted that in his clearest equatorial plate there were 46 chromosomes. Despite this, he later incorrectly concluded that the chromosome number was 48 (Painter, 1923). Thus, when Mittwoch (1952) studied spermatogenesis in a Down syndrome individual and discovered 24 chromosomal masses at diakinesis he did not consider it unusual. It was not until 1956 that improved techniques utilising lung fibroblast cultures were developed and the normal diploid chromosome number was established as 46 (Tjio
and Levan, 1956). This was later confirmed by the observation that in human spermatocytes the haploid number is 23 (Ford and Hamerton, 1956).

Despite the uncertainty about the genetic details, the first suggestion of trisomy as the cause of Down syndrome was made by Bleyer (1934, 1937) and of non-disjunction by Waardenburg (1932). Waardenburg’s idea was disputed because of lack of evidence that non-disjunction was affected by maternal age and because of the discovery of familial cases transmitted by normal carriers which suggested the presence of an abnormal chromosome which occasionally resulted in an individual with unbalanced chromatin content (Penrose, 1939).

It was not until 1959 (Lejeune et al) that it was shown that non-familial Down syndrome individuals had an extra acrocentric chromosome resulting in a diploid chromosome number of 47 in cultured fibroblasts. This was confirmed by Jacobs et al (1959) who described the extra chromosome as 'an acrocentric chromosome in the smallest size range' and suggested that Down syndrome was due to a trisomy of this chromosome. Simultaneously, familial Down syndrome was investigated and in two lineages a normal (46) chromosome complement was discovered. Polani et al (1960) examined a 21 year old mother and found a translocation (15 : 21) due to centric fusion of 2 chromosomes which was present in 3 generations. Fraccaro et al (1960) also reported a chromosomal defect that was identified as either an isochromosome 21 or a translocation (21 : 22).
1.2 Genetics Of Down Syndrome

(Fraser Roberts and Pembrey, 1979: Smith and Berg, 1976)

1.2.1 Non-Familial Down Syndrome

The normal human karyotype consists of 22 pairs of autosomal chromosomes and 1 pair of sex chromosomes (XX / XY) resulting in a diploid number of 46. In non-familial Down syndrome there is an extra chromosome 21. The diploid number in Down syndrome is therefore usually 47 although there have been cases described where Trisomy 21 coexists with other abnormalities e.g. Down syndrome with Klinefelter's syndrome (Trisomy 21, XXY: diploid number, 48) (Ford et al, 1959).

As described above, the incidence of Down syndrome increases in proportion to the age of the mother. Thus the event causing Down syndrome is most likely to occur in the production of the female gamete. Cytogenetic studies concluded that the extra chromosome was maternal in 80.6% and paternal in 19.4% of Down individuals (Hassold and Jacobs, 1984). Recently this has been re-analysed by polymerase chain amplification of DNA polymorphisms in the chromosome 21 of both parents and the affected child (Antonarakis, 1991). Of 200 families examined, 95% of the extra chromosomes were maternal and 5%, paternal.

The process in gamete formation which results in an abnormal number of chromosomes is called non-disjunction (Fig 1.1). The same process may cause many different chromosomal abnormalities (Table 1.1) and has been shown to be maternal age related in the case of autosomes (Patau, 1963) and the sex chromosomes (Court-Brown, 1964).
Disomy 21: Normal Zygote

Trisomy 21: Down Syndrome

Monosomy 21: Lethal

Result of Combination with a Normal Gamete

Figure 1.1. Diagram to show the effect of Non-disjunction during meiosis. The product of the second division is the gamete which may be either nullo-, mono- or diplo- somic for the chromosome in question. If the affected chromosome is number 21, the result of combination with a normal gamete will be either monosomy 21 (Lethal), Disomy 21 (Normal) or Trisomy 21 (Down syndrome).
Table 1.1
Some examples of chromosomal abnormalities due to non-disjunction.

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Chromosome number and abnormality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down syndrome</td>
<td>47, Trisomy 21</td>
</tr>
<tr>
<td>Edwards syndrome</td>
<td>47, Trisomy 18</td>
</tr>
<tr>
<td>Patau syndrome</td>
<td>47, Trisomy 13</td>
</tr>
<tr>
<td>Turner syndrome</td>
<td>45, XO</td>
</tr>
<tr>
<td>Triple-X female</td>
<td>47, XXX</td>
</tr>
<tr>
<td>Multi-X female</td>
<td>48, XXXX</td>
</tr>
<tr>
<td>Klinefelter syndrome</td>
<td>47, XXY</td>
</tr>
<tr>
<td>XYY male</td>
<td>47, XYY</td>
</tr>
</tbody>
</table>
1.2.2 Familial Down Syndrome

In familial Down syndrome there is a fixed genetic defect instead of a trisomy causing Down syndrome. This defect is often a translocation in which chromosome 21 becomes attached to another chromosome by a process called 'centric fusion' (or Robertsonian translocation). One of the earliest descriptions of this type of defect was by Polani et al (1960) who described a Down syndrome child who only had 46 chromosomes. There were 4 in the chromosome 21-22 group (when 5 would be expected in Down syndrome), 5 instead of 6 in the 13-15 group and 15 instead of 14 in the 6-12 group. The chromosome groups were determined by size alone and thus the pattern seen in this patient can be explained by the translocation of chromosome 21 to chromosome 15 which makes it appear to be a member of the 6-12 group. This particular defect (t(15q:21q)) is the commonest centric fusion event found in familial Down syndrome. There are 3 different (Dq:21q) translocations (with chromosomes 13, 14 and 15) and 1 (Gq:21q) translocation (t(22q:21q)) which can occur. By examining the proportions of these in sporadic translocations, the mutation rate for all types of translocation related to Down syndrome has been estimated at $2.71 \times 10^{-5}$ translocations / chromosome 21 / gamete / generation (Polani et al, 1965).

Balanced translocations as described above can be present in apparently normal individuals who act as carriers and if there were no selective force in action there would be 3 equally likely outcomes of pregnancy in carrier families (normal child, carrier child, affected child (Fig 1.2)). This theoretical risk is not observed, (Penrose and Smith, 1966: Hamerton, 1971). If the mother is a heterozygous carrier there is a 90%
Figure 1.2.
Diagram to show the possible gametes deriving from a translocation carrier.
chance that the offspring will be either a normal or a carrier and only a
10% chance of a Down affected individual. If it is the father who is the
heterozygous carrier, there is a 97.4% chance of a normal or a carrier and
only a 2.4% chance of a Down affected child. However, there is a slight
excess of carriers over normals when there is paternal transmission
(Hamerton, 1971). The reasons for the low frequency of Down affected
individuals is not clear but it is likely to be at least partly due to a greater
frequency of early pregnancy loss of affected zygotes (Penrose and Smith,
1960). The cause of the effects in paternal transmission is more difficult
to explain but it has been suggested that extra chromosome 21 material
results in a physical disadvantage to locomotion of the sperm or ability to
fertilise the ovum.

Thus, in conclusion, familial Down syndrome is due to a
chromosomal abnormality which is present in one of the parents
resulting in formation of abnormal gametes and thus a high frequency of
Down affected offspring in the family. This chromosome abnormality
may be present in a carrier form, thus, production of Down affected
offspring is not affected by maternal age. On the converse, non-familial
Down syndrome is due to a failure of normal gametogenesis
(non-disjunction), occurring mainly in the development of the ovum but
occasionally during spermatogenesis. Since most cases are due to non-
disjunction in the ovary, the incidence of non-familial Down syndrome is
strongly associated with maternal age.
1.3 CHARACTERISTICS OF DOWN SYNDROME INDIVIDUALS
(Lyons-Jones, 1988: Cohen and Nadler, 1988)

1.3.1 Major Features

General: Downs individuals are hypotonic with a tendency to keep their mouths partly open and their tongues slightly protruding. They have hyperflexibility of their joints and are relatively short in stature.

CNS: Mental deficiency.

Craniofacial: There is brachycephaly (shortness of the head i.e. cephalic index 81.0 - 85.4: Normal = mesocephalic = 76.0 - 80.9) with a relatively flat occiput and a tendency to a midline parietal hair whorl. In addition, there is microcephaly with upslanting palpebral fissures, hypoplasia of the frontal sinuses, a short highly arched palate, a small nose and a low nasal bridge which results in a tendency to have inner epicanthal folds.

Eyes: The iris is speckled (Brushfield's spots) and peripherally hypoplastic. Fifty nine percent of Down's individuals have fine lens opacities on slit lamp examination and most have refractive errors.

Ears: The ears are small and there is over-folding and angulation of the upper helix with small or absent earlobes.

Teeth: These are hypoplastic and often irregular.

Hands: The metacarpals and phalanges are relatively short. Sixty percent also have hypoplasia of the middle phalanx of the little finger and clinodactyly (an abnormally bent finger). Fifty five percent of Down's
individuals also have a 'simian crease' instead of the usual palmar dermatoglyphic pattern.

**Feet:** There is a wide gap between the 1\textsuperscript{st} and 2\textsuperscript{nd} toes and there is also a plantar crease between these toes.

**Pelvis:** This is hypoplastic with an outward lateral flare and a shallow acetabular angle.

**Heart:** About 40\% of Down's individuals have cardiac anomalies of varying severity. Types (in decreasing order of frequency) include: atrioventricularis communis, ventricular septal defect, patent ductus arteriosus, atrial septal defect, aberrant subclavian artery. (i.e. mostly endocardial cushion defects).

**Skin:** There are loose folds in the posterior neck (nuchal folds). Fifty three percent have Cutis marmorata (a blue/purple mottling of the skin which is either present constantly or only on exposure to cold air) and with time 75\% develop dry hyperkeratotic skin patches.

**Hair:** This is usually fine, soft and sparse. At puberty straight pubic hair develops.

**Genitalia:** The male has a relatively small penis and hypogonadism causing reduced fertility and reduced testosterone production.
1.3.2 Occasional Features

These include: epilepsy (<5%); strabismus (33%); nystagmus (15%); keratoconus (6%); cataracts (1.3%); low placement of the ears; webbed neck; two ossification centres in the manubrium sterni; funnel/pigeon chestedness; tracheo-oesophageal fistula; duodenal atresia; imperforate anus; tetralogy of Fallot; incomplete fusion of the vertebral arches of the lower spine (37%); atlantoaxial instability (12%); abnormal odontoid process (6%); cryptorchism; syndactyly of 2\textsuperscript{nd} and 3\textsuperscript{rd} toes; leukaemia (incidence 1 in 95); thyroid disorders (athyreosis, simple goitre and hyperthyroidism).

1.3.3 Diagnostic Features in the Neonate
(Hall, 1966)

Diagnosis can generally be made shortly after birth and the following features are helpful:

- Hypotonia 80%
- Poor Moro reflex (startle reflex) 85%
- Hyperflexibility of joints 80%
- Excess nuchal skin 80%
- Flat facial profile 90%
- Slanted palpebral fissures 80%
- Anomalous auricles 60%
- Pelvic dysplasia 70%
- Clinodactyly of little finger 60%
- Simian crease 45%
1.3.4 Natural History of Down Syndrome

Muscle tone tends to improve with age in contrast to the rate of
developmental progress which slows with age. Growth is relatively slow
and secondary ossification centres are slow to develop especially during
the first 8 years. During later childhood, osseous maturation is more
normal and final height is usually attained by 15 years. Adolescent
sexual development is usually less complete than normal. Females may
menstrualt and can be fertile but males are usually infertile and have low
serum testosterone concentrations. The major cause of early mortality is
congenital heart defects (44% of those with cardiac anomalies die in
infancy).

1.4 Epidemiology of Down Syndrome


1.4.1 Incidence of Down syndrome

Studies of the epidemiology of Down syndrome are problematical
because it is essential that all of the pregnancy outcomes in the study
group are known. This is particularly difficult if early pregnancy loss is
considered because there is a higher frequency of early foetal loss when
there is a chromosomal abnormality. In a study of chromosomal
abnormality rates at amniocentesis (at 16 weeks gestation) and in live
born infants it has been estimated that only 70% of trisomy 21, 57% of
trisomy 13 and 32% of trisomy 18 survive to term (Hook et al, 1983). The
possibility that foetal loss prior to amniocentesis may be more common
in chromosomally abnormal foetuses must also be considered. The foetal
loss rate described by Hook et al (1983) appeared to be constant across a
maternal age range of 33 - 49 years so it can be assumed that any change
in Down syndrome frequency with maternal age is due to a change in the likelihood of production of an abnormal gamete independently of the foetal loss rate. Thus, foetal loss will be irrelevant when it comes to calculation of Down syndrome risk factors. However, it is extremely relevant to the calculation of screening parameters, particularly when considering detection rates (this will be considered in chapter 9).

Estimations of incidence have been made in two ways: retrospectively from maternity hospital records etc.; and prospectively by examining and recording every birth over a certain period of time. Prospective studies should be more accurate because it is less likely that cases will not be included due to loss of records etc. The incidences from the two types of study will therefore be considered separately.

Most retrospective studies have been performed by reviewing obstetric and paediatric records. Some have in addition examined death certificates and educational records to improve case detection. Table 1.2 summarises a number of retrospective studies which give a combined estimate for the incidence of Down syndrome as 1 in 1080 live births. Prospective studies of Down syndrome incidence are summarised in table 1.3 and give an aggregate estimate for incidence of Down syndrome at 1 in 801 live births. The combined results of these prospective trials are in accordance with the findings of the UK national register of Down syndrome which recorded 1089 cases in 1989 for 687700 births. After correction for terminations due to prenatal diagnosis and natural causes, an estimated birth frequency of 1.4 cases/1000 births for the UK was calculated (Mutton et al, 1991).
### Table 1.2.
A Selection of Retrospective studies of Down syndrome incidence

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Area + period</th>
<th>No. of Births</th>
<th>No. of Affected</th>
<th>Incidence per 1000 births</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malpas (1937)</td>
<td>Liverpool 1923-32</td>
<td>13964</td>
<td>18</td>
<td>1.3</td>
</tr>
<tr>
<td>Beidelman (1945)</td>
<td>Boston 1930-44</td>
<td>12352</td>
<td>42</td>
<td>3.4</td>
</tr>
<tr>
<td>Landtman (1948)</td>
<td>London 1945-48</td>
<td>3593</td>
<td>4</td>
<td>1.1</td>
</tr>
<tr>
<td>Parker (1950)</td>
<td>Washington DC 1939-48</td>
<td>27931</td>
<td>32</td>
<td>1.2</td>
</tr>
<tr>
<td>Stevenson et al (1950)</td>
<td>Boston 1930-41</td>
<td>29024</td>
<td>15</td>
<td>0.5</td>
</tr>
<tr>
<td>Carter &amp; MacCarthy (1951)</td>
<td>London</td>
<td>71521</td>
<td>107</td>
<td>1.5</td>
</tr>
<tr>
<td>Øster (1953)</td>
<td>Denmark 1923-49</td>
<td>84072</td>
<td>71</td>
<td>0.8</td>
</tr>
<tr>
<td>Record &amp; Smith (1955)</td>
<td>Birmingham, UK 1942-52</td>
<td>231619</td>
<td>252</td>
<td>1.1</td>
</tr>
<tr>
<td>Pleydell (1957)</td>
<td>Northamptonshire 1944-55</td>
<td>52729</td>
<td>86</td>
<td>1.6</td>
</tr>
<tr>
<td>Gentry et al (1959)</td>
<td>New York State</td>
<td>1242744</td>
<td>392</td>
<td>0.3</td>
</tr>
<tr>
<td>Babbott &amp; Ingalls (1962)</td>
<td>Pennsylvania 1955-60</td>
<td>25760</td>
<td>26</td>
<td>1.0</td>
</tr>
<tr>
<td>Buchan (1962)</td>
<td>Newcastle, UK 1948-59</td>
<td>61821</td>
<td>92</td>
<td>1.5</td>
</tr>
<tr>
<td>Collmann &amp; Stoller (1962a, 1962b, 1963a)</td>
<td>Victoria, Australia 1942-57</td>
<td>780168</td>
<td>1134</td>
<td>1.5</td>
</tr>
<tr>
<td>Renwick et al (1964)</td>
<td>British Columbia 1955</td>
<td>34138</td>
<td>50</td>
<td>1.5</td>
</tr>
<tr>
<td>Brewis et al (1966)</td>
<td>Carlisle, UK</td>
<td>8528</td>
<td>10</td>
<td>1.2</td>
</tr>
<tr>
<td>Leck (1966)</td>
<td>Birmingham, UK 1950-65</td>
<td>316954</td>
<td>513</td>
<td>1.6</td>
</tr>
<tr>
<td>Halevi (1967)</td>
<td>Israel</td>
<td>90792</td>
<td>97</td>
<td>1.1</td>
</tr>
<tr>
<td>Stark &amp; Mantel (1967)</td>
<td>Lower Michigan 1950-64</td>
<td>2722774</td>
<td>2432</td>
<td>0.9</td>
</tr>
<tr>
<td>Kashgarian &amp; Rendtorff (1969)</td>
<td>Memphis</td>
<td>195412</td>
<td>192</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Total**                  | **6036260**                   | **5607**      | **0.93**        | **(1 in 1080)**          |
### Table 1.3.
A Selection of Prospective studies of Down syndrome incidence

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Area + Period</th>
<th>No. of Births</th>
<th>No. of Affected</th>
<th>Incidence per 1000 births</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenkins (1933)</td>
<td>Chicago 1926-31</td>
<td>3818</td>
<td>6</td>
<td>1.6</td>
</tr>
<tr>
<td>Harris &amp; Steinberg (1954)</td>
<td>Rochester, USA 1944-50</td>
<td>8716</td>
<td>11</td>
<td>1.3</td>
</tr>
<tr>
<td>McIntosh et al (1954)</td>
<td>New York City 1946-53</td>
<td>5964</td>
<td>11</td>
<td>1.8</td>
</tr>
<tr>
<td>Kurland (1958)</td>
<td>Rochester, USA 1945-54</td>
<td>14200</td>
<td>14</td>
<td>1.0</td>
</tr>
<tr>
<td>McDonald (1961)</td>
<td>Watford, UK 1952-55</td>
<td>3179</td>
<td>6</td>
<td>1.9</td>
</tr>
<tr>
<td>Hall (1964)</td>
<td>Sweden 1961-62</td>
<td>25038</td>
<td>38</td>
<td>1.5</td>
</tr>
<tr>
<td>Stevenson et al (1966)</td>
<td>16 countries 1961-64</td>
<td>416695</td>
<td>347</td>
<td>0.8</td>
</tr>
<tr>
<td>Sever et al (1970)</td>
<td>USA</td>
<td>54761</td>
<td>63</td>
<td>1.2</td>
</tr>
<tr>
<td>Wahrman &amp; Fried (1970)</td>
<td>Jerusalem</td>
<td>24245</td>
<td>53</td>
<td>2.2</td>
</tr>
<tr>
<td>Ingalls (1972)</td>
<td>Boston</td>
<td>10493</td>
<td>17</td>
<td>1.6</td>
</tr>
<tr>
<td>Friedrich &amp; Neilson (1973)</td>
<td>Århus</td>
<td>5049</td>
<td>4</td>
<td>0.8</td>
</tr>
<tr>
<td>Harlap (1974)</td>
<td>Jerusalem</td>
<td>42340</td>
<td>103</td>
<td>2.4</td>
</tr>
<tr>
<td>Cuckle et al (1991)</td>
<td>UK 1974-87</td>
<td>8875000</td>
<td>11170</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>10115317</strong></td>
<td><strong>12620</strong></td>
<td><strong>1.24</strong></td>
</tr>
</tbody>
</table>

(1 in 801)
More important than the overall birth incidence is the maternal age related incidence. This has been extensively examined by Cuckle et al (1987) who combined the age related incidences from 8 major surveys (Hook and Chambers, 1977: Hook and Fabia, 1978: Hook and Lindsjö, 1978: Trimble and Baird, 1978: Sutherland et al, 1979: Young et al, 1980: Koulischer and Gillerot, 1980: Huether et al, 1981). They developed a formula for calculating the risk of a mother carrying a Down syndrome affected foetus depending on her age at the expected date of delivery.

There are a number of other congenital conditions which occur. The approximate birth incidences are shown in table 1.4 to allow the incidence of Down syndrome to be put into perspective. It is obvious that Down syndrome provides a large contribution to the total number of congenital abnormalities and that this means that screening for Down syndrome may provide significant benefits.

1.4.2 Survival of Down Syndrome Individuals

In 1929, the life expectancy of Down syndrome individuals was approximately 9 years and by 1947 this had increased to 12 years (Penrose 1932, 1949). In 1954, Brothers and Jago reported that the average age of death of institutionalised Down syndrome cases in Victoria, Australia was 10 years which by 1963 had improved to 18.3 years (Collmann and Stoller, 1963b, 1963c). These survival figures however did not consider the difference between Down syndrome children with or without congenital heart disease. When those affected by heart disease were excluded, the average life expectancy became much greater as the death rate due to heart disease was up to 68% by the age 10 (Fabia and Drolette,
Table 1.4.
Other congenital conditions (birth incidences).

<table>
<thead>
<tr>
<th>Incidence / 1000 live births</th>
<th>Sex Chromosome abnormalities</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Turner's (XO)</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Klinefelter’s (XXY)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Poly-X (female)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Poly-Y (male)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Autosomal Chromosome Abnormalities</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trisomy 21</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>Trisomy 18</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Trisomy 13</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Neural Tube Defects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spina Bifida</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Anencephaly</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Encephalocoele</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Gastrochisis</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Omphalocoele</td>
<td>0.21</td>
</tr>
</tbody>
</table>
1970) (Fig 1.3). In institutionalised Down syndrome individuals the mean age at death was estimated at 35.3 years in England (Richards and Sylvester, 1969) and 30.5 years in Texas (Deaton, 1973). This figure was higher due to a selection bias as only those who survived the initial few years of life became institutionalised. The overall life span of Down syndrome affected individuals has continued to increase and by 1990, the median survival was estimated at 46.5 years (Goldstein and Philip, 1990: Neilsen et al, 1987: Mikkelsen et al, 1977: Dupont et al, 1986). The life table presented by Goldstein and Philip is shown in table 1.5, together with 1983 figures for England and Wales for male normals (Gill et al, 1987).

Other trisomies (13 and 18) have a significantly worse prognosis than Down syndrome with survival to 6 months of approximately 3% (Trisomy 18) and 10% (Trisomy 13) (Goldstein and Neilsen, 1988).

1.5 Health Economics Of Down Syndrome

The incidence of Down syndrome and other trisomies and their survivability is important when the health economics is considered because the costs of a disease depend on its frequency and the duration of time for which an affected individual requires care. The cost of a Down syndrome child can be broken down into 2 major categories: tangible (monetary) and intangible (psychological etc.). Similarly, costs of a screening program for Down syndrome can be separated into these categories. Examples of these costs and benefits are shown in table 1.6.

By totalling costs and benefits it is possible to determine whether a procedure is financially beneficial or not. This accounting-based method
Figure 1.3. Survival of Down syndrome children up to age 10 years depending on presence or absence of congenital heart disease (Fabia and Drolette, 1970).
Table 1.5.
A Life table for Down syndrome individuals. Value = probability of individual being alive at each age.
(Goldstein and Philip, 1990: Gill et al, 1987).

<table>
<thead>
<tr>
<th>Age</th>
<th>Down syndrome affected</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>0.8543</td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>0.7877</td>
<td>0.989</td>
</tr>
<tr>
<td>2-3</td>
<td>0.7819</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>0.7492</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>0.7371</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>0.7371</td>
<td>0.987</td>
</tr>
<tr>
<td>6-7</td>
<td>0.7211</td>
<td></td>
</tr>
<tr>
<td>7-8</td>
<td>0.7051</td>
<td></td>
</tr>
<tr>
<td>8-9</td>
<td>0.6891</td>
<td></td>
</tr>
<tr>
<td>9-10</td>
<td>0.6731</td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>0.6649</td>
<td>0.986</td>
</tr>
<tr>
<td>15-19</td>
<td>0.6551</td>
<td>0.984</td>
</tr>
<tr>
<td>20-24</td>
<td>0.6325</td>
<td>0.980</td>
</tr>
<tr>
<td>25-29</td>
<td>0.6137</td>
<td>0.976</td>
</tr>
<tr>
<td>30-34</td>
<td>0.5954</td>
<td></td>
</tr>
<tr>
<td>35-39</td>
<td>0.5775</td>
<td>0.967</td>
</tr>
<tr>
<td>40-44</td>
<td>0.5595</td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>0.5189</td>
<td>0.951</td>
</tr>
<tr>
<td>50-54</td>
<td>0.4582</td>
<td></td>
</tr>
<tr>
<td>55-59</td>
<td>0.3269</td>
<td>0.897</td>
</tr>
<tr>
<td>60-64</td>
<td>0.1607</td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>0.0794</td>
<td>0.740</td>
</tr>
<tr>
<td>70-74</td>
<td>0.0055</td>
<td></td>
</tr>
<tr>
<td>78-79</td>
<td></td>
<td>0.400</td>
</tr>
</tbody>
</table>
Table 1.6.
Examples of costs and benefits in Down screening (Gill et al, 1987).

**Costs of Down Child**

* Tangible

Lost parental output
Loss of expected output from child
Consumption of goods and services
Use of health care services
Special education facilities
Fostering and adoption

* Intangible

**Costs of Screening**

* Tangible

Counselling
Sample collection and analysis
Lost production whilst at clinic
Counselling after diagnosis
Cost of termination (TOP)
Lost production during TOP
Travel to hospital

* Intangible

Psychological: false security
Pain/anxiety/discomfort of tests
Anxiety waiting for results
Psychological cost of TOP

**Benefits of Screening**

* Tangible

Saving of resources which would have been spent on Down child
Retention of parental output
Saving of maternity leave costs
Saving of costs of birth

* Intangible

Avoiding handicapped child
Avoiding -ve effect on siblings
Avoiding -ve effect on parents (divorce etc.,)
Avoids parental worry when they become old and are unable to care for Down child
results in a benefit : cost ratio which must be greater than 1.5 for the procedure to considered worthwhile. However, there are problems in this type of evaluation: it is only really possible to consider those costs and benefits which can have a monetary value attached to them. Thus, the psychological reassurance which may be provided to women receiving a negative result and the extra anxiety caused to those women given a positive result cannot be quantified or included in the benefit : cost ratio.

Goldstein and Philip (1991) examined the benefit : cost figures for an amniocentesis-only based screening program for Down syndrome, trisomy 13 and trisomy 18. They found that depending on the discount rate (inflation cost modifier), there was a benefit in performing amniocentesis particularly in older women (>30 years).

In 1987, Gill et al appraised the benefits of screening for Down syndrome using a combination of maternal age and α-fetoprotein and found a benefit : cost ratio of 23.6 for maximal uptake (100% of women accepting screening) with a replacement rate of 0 (i.e. no further attempts at reproduction) and a discount rate of 5% or 12.2 with a more realistic uptake rate (50%), replacement rate of 100% and discount rate of 7%. Extension of the Down screen to include HCG and uE₃, was examined by Sheldon and Simpson (1991) but they used the measure of cost / birth of Down syndrome baby avoided which makes comparison difficult. The average cost per avoided birth in 1991 was £29,341 (for screening with 80% acceptance of serum testing and 75% acceptance of amniocentesis). The lifetime costs of a Down syndrome baby in 1987 were estimated to be £126,590 based on a 5% discount rate. Assuming 5% per annum inflation
this is equivalent to £153,870 in 1991 and gives a benefit : cost ratio of 5.2. This still reflects a significant benefit which implies that increasing the number of analytes in the screen is financially viable as well as benefiting patients.
chapter 2:

**INTRODUCTION TO DOWN SCREENING**

2.1 **GENERAL PRINCIPLES OF SCREENING**

2.1.1 Definitions
2.1.2 Ethical Considerations

2.2 **DEVELOPMENT OF DOWN SYNDROME SCREENING**

2.3 **THE MEANING OF RISK**
2.1 GENERAL PRINCIPLES OF SCREENING  

2.1.1 Definitions

Screening: screening can be defined as the presumptive identification of unrecognised disease or defect by the application of tests, examinations or other procedures which can be applied rapidly. Thus, screening tests sort out those apparently well people who may have a condition from those who probably do not. i.e. screening tests are not diagnostic and if the screening test gives a positive result, a diagnostic test is required.

screened individuals can be divided into four groups:

True positives (TP): those individuals who have a positive screening test result and who have the condition screened for.

True negatives (TN): those individuals who have a negative screening test result and who do not have the condition screened for.

False positives (FP): those individuals who have a positive screening test result but who do not have the condition screened for.

False negatives (FN): those individuals who have a negative screening test result but do have the condition screened for.

These groupings are useful because they allow the calculation of indices which can be used to assess the effectiveness of a screening program.
The derived indices are:

**Sensitivity**: (equivalent to detection rate): sensitivity is defined as the ability of a test to give a positive result in people who have the condition being screened for (i.e. positivity in disease).

\[
\text{sensitivity} = \frac{TP}{TP + FN} \times 100\%
\]

**Specificity**: (equivalent to false positive rate): specificity is defined as the ability of a test to give a negative result in people who do not have the condition being screened for (i.e. negativity in health).

\[
\text{specificity} = \frac{TN}{TN + FP} \times 100\%
\]

**Predictive value**: predictive value gives an idea of how effective a test is in the clinical environment. The predictive value of a positive result is the likelihood that a positive result means that the condition is present and similarly, the predictive value of a negative result gives the likelihood that a negative result indicates the condition is absent. The predictive value is probably the most important test characteristic because it varies according to the frequency of the screened condition.

\[
\text{Predictive value of a +ve result} = \frac{TP}{TP + FP} \times 100\%
\]

\[
\text{Predictive value of a -ve result} = \frac{TN}{TN + FN} \times 100\%
\]
**Efficiency:** efficiency relates to the overall performance of a screening test.

\[
\text{efficiency} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%
\]

**ROC plot:** When the above indices are examined critically, it is evident that as sensitivity increases, specificity decreases. This relationship can be represented graphically by plotting sensitivity (y axis) against specificity (x axis). This type of graph is called a receiver operator curve after its original use to describe the decision making behaviour of radar operators. Examination of the curve allows the optimum screening cut-off to be determined such that the detection rate is maximised for a tolerable false positive rate.

The meaning of these indices is:

A sensitive test is one which rarely misses disease and a specific test is one which rarely causes false alarms. In general practice the most important feature of a screening test is its specificity because the incidence of any particular disease is generally low. In a hospital setting, sensitivity is more important because the population under scrutiny is already selected as being 'ill' and the prevalence of disease is thus greater (Mathers and Hodgkinson, 1989). i.e. the predictive value of the parameters is affected significantly by the prevalence of the condition being screened for.
2.1.2 Ethical Considerations

It is now acknowledged that screening has the potential to do more harm than good (Mant and Fowler, 1990). In the 1950's screening for lung cancer by mass radiography and sputum cytology was advocated and after 5 years of follow up it appeared to be successful with 35% survival in the screened group, twice that of controls. However, by 10 years, the cumulative mortality in both groups was identical (Fontana, 1985). This experience should be remembered by any person who is considering initiating a screening program.

In practice screening is often promoted before due consideration of its efficacy or feasibility. Current screening programs include: screening for cervical neoplasia and breast cancer (for which there is evidence of efficacy (Roebuck, 1986: Day, 1989) although in the case of breast cancer screening this is still disputed (Anderson et al, 1988)); screening for cardiovascular disease (for which efficacy has not yet been proven); screening and treatment of hypertension (which has been shown to decrease the risk of myocardial infarction and stroke (Wilcox et al, 1986); screening by urinalysis, faecal occult blood and breast examination (all of which have uncertain efficacy (Simon, 1985: Zilva, 1985: Mant, 1989)).

It must be noted that efficacy in clinical trials does not equate to efficacy in clinical practice. The minimum criteria that a screening procedure must fulfil before it becomes of value in general practice and therefore becomes ethically acceptable, have been defined (Mant and Fowler, 1990). These are:
1) Can we offer effective treatment to patients if the screened condition is present?

2) How many positive tests will prove to be false alarms (and is this acceptable)?

3) How many patients will need follow up over the next 5 years (and can the workload be sustained)?

4) How can routine audit of the quality of the test, the intervention and its follow up be performed?

In the case of Down syndrome screening, question 3 is not relevant because the screen is a one-off event for which long term follow up is inappropriate. The remaining questions are all very relevant. Question 1 can be answered easily because in the event of a positive screening test, a diagnostic test (amniocentesis and karyotyping of cultured cells) can be performed. This allows confirmation of an affected fetus and either termination or provision of counselling and social service assistance in the event of a decision to continue the pregnancy. The answers to questions 2 and 4 will be dealt with in greater detail in chapters 8 and 9 respectively.

Standards of medical practice have traditionally been set by physicians on the basis of their judgement of what constitutes good patient care but other factors such as cost containment, profit motive, government regulation, fear of medical malpractice litigation and 'media hype' frequently influence practice more than professional standards.
(Elias and Annas, 1987). In certain countries the ability of a patient to pay for a test may also be a deciding factor. Furthermore, in an antenatal screening program such as Down syndrome screening there are a number of extra considerations which must be examined (Elias and Annas, 1987: Chamberlain, 1991: Harper, 1992). Based on these papers, a set of ethical guidelines for amniocentesis can be compiled.

1) Diagnosis should be made as early as possible in pregnancy (pre-conception is ideal but rarely possible).

2) Counselling should be provided before screening is performed. i.e. testing should not be routinely performed; rather, it should be routinely available allowing the patient to opt out of the screen if they so wish. Initial pre-screening counselling may be provided as an information booklet or a video presentation, provided that verbal information is available if required by the patient.

3) Counselling should be non-directive, i.e. the family should be allowed to make their own decision based on information provided rather than on emotive phraseology. Thus, results should be provided as a risk value with a comment indicating whether the risk is higher or lower than the 'usual' risk cut-off in the screening program. Phrases such as 'Screen positive' and 'Screen negative' are not suitable and should be avoided.

4) In the event of a 'positive' screen, no pressure to opt for termination should be applied and arrangements should be
made to ensure that the baby is born in the most appropriate
centre to ensure survival if termination. Furthermore,
community services should be alerted in the event of
continuation of pregnancy.

5) Screening should not be carried out until all relevant staff have
been sufficiently trained and protocols established for
management of patients with positive results.

6) Screening must be carried out openly to avoid the excesses of
the past (the eugenics and Nazi movements both believed
that subordination of individual decisions for the propagation
of broader population goals were justified and that the
resultant over-riding of individual rights would benefit the
population in the long term). Furthermore, informed public
debate should be encouraged because it educates the
population and allows them to make reasoned choices.

7) Consideration must be given to patients with different ethnic,
cultural and religious backgrounds to prevent them being
pressurised into being screened without their full consent.

8) The risk to unaffected pregnancies must be considered before a
decision on the acceptable frequency of amniocentesis can be
made. i.e. knowledge of the complication rate of amniocentesis
is vital because it would be wrong to harm a large number of
normal fetuses in order to screen for a small number of
affected pregnancies.
These ethical considerations tend to be applied without any form of enforcement because they are logical but we must take care to avoid any future ethical problems. Therefore, evaluation of the ethical background to screening should be included in continuous clinical audit of antenatal screening for Down syndrome.

2.2 Development of Down Syndrome Screening

One of the earliest proposals for Down syndrome screening was made in 1973 by Stein et al who suggested that a phased introduction of amniocentesis for all women was feasible and could potentially prevent most Down syndrome births if a sufficient acceptance rate were achieved. This scheme was never implemented in full, but amniocentesis for women aged 35 years and over did become an accepted norm. For this selection criterion, Stein et al predicted that a 90% acceptance rate would result in detection of 30% of all cases of Down syndrome. This acceptance rate however was not achieved in practice; Quarrell et al (1987) reviewed uptake figures for the University Hospital of Wales and found that for women offered amniocentesis on the basis of age alone the uptake rates were:

- Age 35 - 36 years: Uptake was 16%.
- Age 37 - 39 years: Uptake was 35%.
- Age >= 40 years: Uptake was 46%.

These figures are comparable with uptake rates reported by Youings et al (1991). Thus, amniocentesis programs based on maternal age alone have only limited success in identifying Down syndrome.
affected pregnancies.

Almost simultaneously with the suggestion of Stein et al, Brock and Sutcliffe (1972) showed that antenatal detection of another common fetal malformation was possible. They showed that open neural tube defects (NTD) caused increased maternal serum concentrations of α-fetoprotein (AFP) and this led to the setting up of a collaborative study in the UK which started in January 1975. The results were published in 1977 (UK collaborative study on AFP in relation to NTD) and established the expression of AFP results as multiples of the median (MoM) to allow for the increasing concentration of maternal serum AFP as gestation progressed. It was shown that the optimum sampling time (singleton pregnancies) was 16 - 18 weeks of gestation and that with a 2.5 MoM cut-off, there was a detection rate of 88% for anencephaly and 79% for spina bifida with a 3% false positive rate giving a predictive value of a positive result of approximately 10%.

By 1980, AFP screening for NTD was a well proven technique but caution continued to be urged because the prevalence of NTD was low and there was worry that poor quality control of AFP assays could increase the false positive rate and potentially result in the termination of normal pregnancies (either as a complication of amniocentesis or as a result of errors in analysing the amniotic fluid sample (Davidson and Walker, 1980: Mennuti, 1980). The acceptance of AFP screening for NTD detection was instrumental in the development of Down syndrome screening because it allowed collection of the data which showed that AFP concentrations were decreased in pregnancies with chromosomally abnormal fetuses.
In 1982, a 28 year old primigravid nurse delivered an infant with trisomy 18. She had earlier been screened for NTD by serum AFP testing and the result had been described as 'below the minimum sensitivity' of the assay. She had so nagged her physicians about the significance of this that they decided to examine the AFP results of other pregnancies resulting in chromosomally abnormal infants (Macri J: personal communication). They identified 44 cases of fetal autosomal trisomy from a series of genetic amniocenteses and and found that the median concentration of AFP in the trisomy affected pregnancies was significantly lower than in 'normal' pregnancies (Merkatz et al, 1984). Four weeks later, a second study of 61 trisomic pregnancies was published (Cuckle et al, 1984) which showed that the median AFP MoM in Down syndrome was 0.72 with a similar variance to that of 'normal' pregnancies which had a median MoM of 1.0. Both groups suggested that these low concentrations of AFP could be used in conjunction with maternal age to improve screening for Down syndrome in younger women, although the detection rate would still be rather low. The following year, a meta-analysis of trisomy 21 pregnancies confirmed that there was a significant decrease in maternal serum AFP (median MoM = 0.79) but emphasised caution due to the low predictive value of AFP screening (Spencer and Carpenter, 1985).

The above studies used simple MoM cut-off values for different age groups to determine which women should be offered amniocentesis on the basis of a low AFP result. Conversion of MoM results to a 'risk' was suggested by Baumgarten et al (1985a) who used a linear transformation (see chapter 3 for risk calculation methods) and showed that using a risk cut off of 1:270, about 5% of women under 35 could be identified, amongst
whom would be 25 - 33% of Down syndrome affected pregnancies (DiMaio et al., 1987). Palomaki and Haddow (1987) further improved the risk calculation method by using a Gaussian distribution function to model the situation and allow determination of whether the AFP MoM is more likely to result from a normal or a Down syndrome pregnancy population. This was achieved by calculation of a likelihood ratio and modifying the maternal age related risk (see chapter 3). Three months later Cuckle et al. (1987) published a similar method but included an equation which could be used to calculate maternal age specific risks at the expected date of delivery. This formula (see chapter 3) is immensely useful because it simplifies computerisation of risk calculation by removing the need for large tables of risk constants. Table 2.1 shows results from a selection of studies of the effectiveness of AFP screening either using MoM cut-offs or in combination with age as 'risk screening'.

Shortly after the description of screening using AFP and age, several other markers were found to be useful. These will be dealt with in order of publication:

Firstly, human chorionic gonadotropin (HCG). Bogart et al. (1987) showed that the concentration of HCG in maternal serum was significantly raised and that if $\beta$-HCG (intact HCG plus the free $\beta$ subunit of HCG) and $\alpha$-HCG (free $\alpha$ subunit) results without correction for gestational age in patients at 18 weeks of gestation were combined without considering maternal age then there was a detection rate of 76% with a false positive rate of 4.1%. Combination of increased HCG with decreased AFP by ratio was shown to be effective with a detection rate of 62% and a false positive rate of 10% or a detection rate of 38% with a false
Table 2.1. Effectiveness of AFP screening for chromosomal abnormalities.

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Method</th>
<th>No. of affected pregnancies</th>
<th>Detection Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merkatz et al (1984)</td>
<td>MoM &lt; 0.4</td>
<td>41</td>
<td>25%</td>
<td>5.3%</td>
</tr>
<tr>
<td>Cuckle et al (1984)</td>
<td>MoM &lt; 0.5</td>
<td>61</td>
<td>21%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Spencer and Carpenter (1985)</td>
<td>MoM &lt; 0.5</td>
<td>27</td>
<td>17%</td>
<td>8.9%</td>
</tr>
<tr>
<td>DiMaio et al (1987)</td>
<td>Risk &gt;1:270</td>
<td>36</td>
<td>36%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Bogart et al (1987)</td>
<td></td>
<td>25</td>
<td>24%</td>
<td>6.8%</td>
</tr>
<tr>
<td>Aitken et al (1988)</td>
<td></td>
<td>142</td>
<td>37%</td>
<td>6.6%</td>
</tr>
<tr>
<td>Wald et al (1988a)</td>
<td></td>
<td>77</td>
<td>35%</td>
<td>5.3%</td>
</tr>
<tr>
<td>Suchy and Yeager (1990)</td>
<td></td>
<td>16</td>
<td>25%</td>
<td>4.8%</td>
</tr>
<tr>
<td>Zeitune et al (1991)</td>
<td>Risk &gt; 1:280</td>
<td>142</td>
<td>37%</td>
<td>6.6%</td>
</tr>
</tbody>
</table>
positive rate of 2% (Arab et al, 1988). White et al (1989) suggested that the AFP / HCG ratio may be evaluated using a Gaussian frequency
distribution method similar to that described by Palomaki and Haddow
(1987) for AFP only screening. This was better than a simple ratio cut-off
method (detection rate = 100% with false positive rate = 11%; detection
rate = 73% with false positive rate = 4%). Results from a selection of
studies of HCG are shown in table 2.2. These show that combination of
AFP and HCG results to identify Down syndrome pregnancies is superior
to screening with either marker alone.

Secondly, unconjugated estriol (uE3). Low maternal serum uE3 in
the second trimester was first described in Down syndrome pregnancies
by Canick et al (1988) who showed that the median MoM in affected
pregnancies was 0.79. On its own, uE3 offers a 35% detection rate with a
3.8% false positive result. In combination with AFP there is a 45%
detection rate with a 5.3% false positive rate (Wald et al, 1988b). It has
been claimed that combination of AFP, HCG and uE3 further improves
screening (detection rate = 60% with false positive rate = 4.7%) (Wald et al,
1988a). This is said to be an improvement in detection over screening
with a combination of 2 analytes and the superiority of the 'triple test' has
been supported by other workers. Nørgaard-Pedersen et al (1991) showed
an increase in detection rate from 48% to 52% with a constant false
positive rate of 5%; MacDonald et al (1991) showed an increase in
detection rate from 56% to 60% for an increase in false positive rate from
4.1% to 4.6%. The 'triple test' is however rejected by other workers who
believe that unconjugated estriol is of no value whatsoever (Macri, 1990a,
b, c, d, 1991). There are a number of possible explanations for this: the
unconjugated estriol assay may not be sufficiently sensitive or precise;
Table 2.2  
Effectiveness of AFP / HCG screening for chromosomal abnormalities.  

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Number of Affected</th>
<th>Analyte(s)</th>
<th>Detection Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bogart et al (1987)</td>
<td>25</td>
<td>AFP only</td>
<td>24%</td>
<td>4 - 8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HCG only</td>
<td>56%</td>
<td>1.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-HCG + α-HCG</td>
<td>76%</td>
<td>4.1%</td>
</tr>
<tr>
<td>Arab et al (1988)</td>
<td>29</td>
<td>AFP/HCG ratio</td>
<td>62%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>38%</td>
<td>2%</td>
</tr>
<tr>
<td>Wald et al (1988a)</td>
<td>77</td>
<td>AFP + HCG</td>
<td>55%</td>
<td>5.0%</td>
</tr>
<tr>
<td>White et al (1989)</td>
<td>15</td>
<td>AFP + HCG</td>
<td>100%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>73%</td>
<td>4%</td>
</tr>
<tr>
<td>Bogart et al (1989)</td>
<td>20</td>
<td>β-HCG &gt; 2.5 MoM</td>
<td>80%</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-HCG &gt; 2.0 MoM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petrocik et al (1989)</td>
<td>38</td>
<td>HCG &gt; 2.0 MoM</td>
<td>63%</td>
<td>14.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HCG &gt; 2.0 MoM</td>
<td>37%</td>
<td>5.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AFP + HCG + Age</td>
<td>68%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Suchy and Yeager (1990)</td>
<td>16</td>
<td>AFP + HCG + Age</td>
<td>63%</td>
<td>4.7%</td>
</tr>
<tr>
<td>Bartels et al (1990)</td>
<td>42</td>
<td>MoM cut offs</td>
<td>44%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Crossley et al (1991)</td>
<td>78</td>
<td>AFP/HCG ratio + Age</td>
<td>57%</td>
<td>5%</td>
</tr>
<tr>
<td>Miller et al (1991)</td>
<td>19</td>
<td>AFP + HCG + Age</td>
<td>74%</td>
<td>?</td>
</tr>
<tr>
<td>Bogart et al (1991)</td>
<td>11</td>
<td></td>
<td>65%</td>
<td>7.9%</td>
</tr>
</tbody>
</table>
addition of extra analytes may increase imprecision to unacceptable levels; or the detection and false positive rates assessed by the different groups may not be comparable. These factors will be discussed in chapters 7 and 8.

In addition to AFP, HCG and uE₃, a number of other minor markers have been reported. These have included: pregnancy specific β₁-glycoprotein (SP-1) (Bartels and Lindeman, 1988: Bartels et al, 1990: Petrocik et al, 1990) which is significantly elevated to a median MoM of 2.1 in affected pregnancies (this marker was shown to be of little use by Wald et al (1989) who used it as a fourth marker and by Graham et al, (1992) who tried it as a replacement for HCG); Ca-125 which has been reported to be elevated in affected pregnancies by one group (Høgdall et al, 1992) but not elevated by another (Davies et al, 1991); urea resistant neutrophil alkaline phosphatase which is elevated in Down syndrome affected pregnancies (Grozdea et al, 1990: Cuckle et al, 1990); immuno-reactive inhibin (Van Lith et al, 1992) although this is strongly correlated with HCG and may therefore be of little use (Spencer et al, 1993; pregnancy associated plasma protein A (PAPP-A) which has been shown to be low in Down syndrome pregnancies by Wald et al, (1992c, 1992d) and Brambati et al, (1991) but to be unaffected by Knight et al, (1993); and the free β subunit of HCG which is advocated by 2 groups (Macri et al, 1990e: Spencer, 1991: Spencer and Macri, 1992) but criticised by others (Thomas et al, 1990: Knight and Cole, 1991). Claims for free β HCG will be examined in chapter 8.

Biochemical parameters are not the only markers which may be considered for Down syndrome screening in early pregnancy:
Ultrasound derived measurements have also been suggested to be of value. Three main ultrasound signs have been described. These are altered head measurements, increased nuchal fold thickness and alteration of bone growth patterns. Fetal cephalometry was dismissed at about the same time as biochemical screening became possible (Perry et al., 1984).

Nuchal fold thickness was the first sign described that gave promise of effectiveness. This sign involves measurement of the thickness of the skin fold of the posterior neck of the fetus. It was first described in a retrospective study of 904 fetuses (Benacerraf et al., 1985a) which was later expanded to 1704 fetuses (Benacerraf et al., 1985b). Eleven of these fetuses had Down syndrome of which 5 (45%) had a nuchal fold thicker than 7 mm. There was only 1 false positive case (0.06%). A further expansion of this series to 3825 fetuses showed detection of 9 of 21 (42%) (Benacerraf et al., 1987a). This sign was dismissed however by Toi et al. (1987) who found a false positive rate of 21% and that the sign could be produced artefactually by slanting the ultrasound probe.

The next development was the combination of a nuchal fold thickness greater than 6 mm with either an increased biparietal diameter (BPD) : femur length (FL) ratio (Lockwood et al., 1987) or an actual : expected FL ratio < 0.91 (Benacerraf et al., 1987b). The combination of nuchal fold thickness with BPD : FL ratio could identify 50 - 70% of affected fetuses with a false positive rate of 6% and the combination of nuchal fold thickness with actual : expected FL ratio identified 75% of affected fetuses with a false positive rate of 2%. This was not confirmed by other workers who found a 42% detection rate with a 12% false
positive rate for nuchal fold thickness plus actual : expected FL ratio (Perrella et al, 1988) and an 8% detection rate with a 7% false positive rate for nuchal fold thickness plus BPD : FL ratio (Ginsberg et al, 1990).

A further series of reports showed that nuchal fold thickness, actual : expected FL and BPD : FL ratios were all poor markers for Down syndrome with detection poorer than that achieved by maternal age screening (FitzSimmons et al, 1989: Marquette et al, 1990: Lockwood et al, 1991). Since femoral markers had been shown to be effectively useless, actual : expected humeral length ratios were considered and shown to be possibly of use (Benacerraf et al, 1991: Rodis et al, 1991). FitzSimmons et al (1989) also examined humeral length and found it to be of interest but could not demonstrate any significant shortening. The future development of ultrasonic screening is beyond the scope of this thesis.

2.3 The Meaning of Risk

The word risk is probably derived from the Greek word 'rhiza' which related to the hazards of sailing too near to cliffs, contrary winds and swirling tides. The idea of risk management was well understood by the Greeks and Romans who identified many common hazards and potentially effective ways of minimising their capacity to cause harm. Hippocrates was one of the earliest writers on this topic when in 400 BC he recognised that sharp objects cause injury by concentrating the forces exerted upon them by the human body. A later writer, Vituvious, also before the birth of Christ, recognised the ill effects of the fumes from lead smelting and concluded that water should not be carried in lead pipes. Traffic safety was also of concern in ancient Rome where wheeled
vehicles and pedestrians were kept segregated during the daytime. Following the decline of the Graeco-Roman civilisation, risks tended to be misunderstood and supernatural forces were often blamed for large and small disasters, crop failures, floods, droughts, diseases etc. Hundreds of thousands of people were, after careful analysis by church dignataries, put to death for practising witchcraft. Eventually, logical thought prevailed and the concept of probability and risk became understood and mathematical theories of the nature of risk were developed.

Prior to the development of the serum Down syndrome screening test, amniocentesis was offered to women aged 35 years and over. This equates to a risk of Down syndrome of greater than 1 in 385. Where serum screening is performed, the risk cut off used is often 'greater than or equal to 1 in 300'. But what does this mean?

Mathematically a risk has a precise definition. For a risk of 1 : n there is 1 chance in n trials of an event occurring. Thus, as n (the number of trials) decreases, the chance (or risk) of an event occurring increases. To clarify this, a risk of 1 : 200 is greater than a risk of 1 : 300, i.e. the larger the value of n, the less chance there is of the event happening.

In human activity however, risk is not a clear cut concept. Few people would back a racehorse with a 1 in 300 chance of winning because the chance of the event is too small but similarly few people would be prepared to fly in an aeroplane which had a 1 in 300 chance of crashing because the risk is too great. This analogy shows that risk is perceived differently depending on the severity of the outcome event. If the outcome event is fixed, e.g. the possibility of a defined financial loss, then
the likelihood of any particular individual taking the risk depends on two factors; their perception of the risk of losing and whether they consider they can afford to take the loss if it occurs. These decisions are inevitably defined by the personality of the 'risk-taker' and defy explanation.

A second concept to consider is inconsistency. A good example of this is that there is a greater risk of death in a bus whilst travelling to an airport than there is of being killed by a terrorist attack whilst flying, yet surveys show that passengers are more worried about the risk of terrorism than the risk of road transport. This inconsistency of evaluation of different risks within an individual is also exhibited between individuals in that there are two basic character types; those who are prepared to accept risk (e.g. rock climbers, hang glider pilots etc.) and those who attempt to avoid risks at all times. The type of risk which is considered unacceptable may also be inconsistent, i.e. those people who are prepared to take physical risks such as rock climbing usually would not consider taking the risk of smoking whereas many smokers do not worry about the risk of smoking but consider climbing too dangerous.

A third important concept of risk is dread. This may also be expressed as 'familiarity breeds contempt' e.g. workers in hazardous occupations often fail to take precautions because they are so used to the danger that they no longer fear it. Similarly, there is a tendency to fear unknown risks more than risks which are known based on dread. For example, the risk of skateboarding is slightly greater than that due to nuclear weapons but people tend to be more worried about nuclear weapons; and the risk due to motor vehicles is approximately the same
Figure 2.1.
Diagram showing the concept of risk and dread (Stewart, 1990). The graph is derived from the Royal Society report on risk assessment (1983).
as that from food irradiation but food irradiation is a very emotive issue. A graph which demonstrates the concept of dread is shown in figure 2.1. The dread factor can partly explain some of the inconsistencies of risk behaviour such as the bus and terrorist example above.

The fourth concept which should be examined is the 'Law of Averages'. This is epitomised by the belief that lightning never strikes in the same place twice and that during wartime, the safest place for a soldier to shelter is an a bomb crater. Both of these examples show that the 'Law of Averages' is fallacious because neither lightning nor a bomb 'knows' where a previous strike occurred and therefore, since avoidance of a previous strike site cannot occur, the probability of an event is at least the same for every trial. Furthermore, in the case of lightning strikes, lightning tends to strike prominent sites because they accumulate electrostatic potentials better. Thus, lightning is actually more likely to strike a site for a second time.

Finally, the additive nature of risk. According to the old Central Electricity Generating Board, the probability of a catastrophic accident (equal to or greater than Chernobyl in severity) at any one nuclear power station in the UK is 1 accident every 10,000 years. This sounds to be a very low risk but there are about 40 nuclear stations in the UK so the cumulative risk is 40 chances every 10,000 years or 1 in 25 years.

The net result of the above points is that risk is an extremely difficult concept to express because different people perceive similar risks differently and decide whether it is acceptable or unacceptable based on their own internal value judgements. Furthermore, two events with a
similar risk may also be perceived differently. Thus, risk perception is intensely personal and it is impossible to define what risk means to any one individual. The only way in which risk can be discussed therefore is to compare two events which are nearly identical and which have approximately equal risks.

In the case of Down syndrome screening, the risks which are generally compared are the risk of Down syndrome and the risk of fetal loss due to amniocentesis. At the University Hospital of Wales it is estimated that the probability of fetal loss is 0.4% (Dawson A.: Personal Communication) and the risk cut-off used to determine which women should be offered amniocentesis is 1: 300 (0.33%). These two risks are very similar and it is this similarity which underpins Down syndrome screening, i.e. since the risk of fetal loss due to screening is equal to the risk of Down syndrome, the benefits of screening are equal to the hazards and therefore screening is unlikely to cause more harm than good (see chapter 2: section 2.1.2; question 2).
3.1 **BIOCHEMICAL ASSAY METHODS**

3.1.1 Royal Gwent Assays
   3.1.1.1 AFP
      3.1.1.1.1 AFP-free serum preparation
   3.1.1.2 total HCG
3.1.2 UHW Assays
   3.1.2.1 Magnetic Solid Phase Preparation
   3.1.2.2 AFP
   3.1.2.3 total HCG
3.1.3 Amersham Assays
   3.1.3.1 AFP
   3.1.3.2 total HCG
   3.1.3.3 unconjugated oestriol 2nd trimester assay
   3.1.3.4 unconjugated oestriol 3rd trimester assay
   3.1.3.5 unconjugated oestriol 3rd trimester assay, Canick / Wald’s modification
3.1.4 CIS free β HCG assay

3.2 **MATHEMATICAL METHODS**

3.2.1 Age Related Risk Calculation
3.2.2 Calculation of Multiples of the Median (MoM)
3.2.3 Weight Correction of MoMs
3.2.4 Risk Calculation Methods
   3.2.4.1 Linear algorithm
   3.2.4.2 Discriminant Function Method
   3.2.4.3 Likelihood Ratio Method
3.2.5 Statistical and Matrix Functions
   3.2.5.1 Gaussian Distribution Formulae
   3.2.5.2 Matrix Mathematics
      3.2.5.2.1 Matrix Notation and Orientation
      3.2.5.2.2 Matrix Addition and Subtraction
      3.2.5.2.3 Matrix Multiplication
      3.2.5.2.4 Determinants
      3.2.5.2.5 Matrix Inversion

3.3 **SIMULATION METHODS**
3.3.1 Random Normal Deviate Method
3.3.2 Cholesky Root Method for Correlated Variables

3.4 Statistical Methods

3.4.1 Deming's Regression
3.4.2 Skewness and Kurtosis
3.4.3 Confidence Intervals for Medians
3.4.4 Kolgomorov-Smirnov type tests

3.5 Ultrasound Formulae
3.1 BIOCHEMICAL ASSAY METHODS

3.1.1 Royal Gwent Hospital Laboratory Methods

The laboratory at the Royal Gwent Hospital (RGH) measures AFP and total HCG concentrations using in-house assays. In later chapters these will be referred to as RGH-AFP and RGH-HCG respectively.

3.1.1.1 Serum $\alpha$-fetoprotein assay (RGH-AFP)

**Principle:** Radioimmunoassay (RIA).

**Reagents:**

1) Assay buffer: 50 mmol/L phosphate buffer containing 0.2% (w/v) bovine serum albumin, pH 7.4.

   Dissolve 6.3g anhydrous disodium hydrogen orthophosphate (Analar), 1.75g sodium dihydrogen orthophosphate dihydrate (Analar), 2g bovine serum albumin (Armour), and 2g sodium azide in 800mL distilled water. Adjust pH to 7.4 and make up to 1L with distilled water. Store at 4°C. Reagents from Merck Ltd, Poole, Dorset.

2) Tracer ($^{125}$I-AFP):

   a) Stock solution: 1mL obtained monthly from Dr. J.L.Young, RIA laboratory, Ninewells Hospital, Dundee. This contains 2µg of labelled AFP with 25µCi/µg.
   b) Working solution: To 10mL assay buffer add 50µL of stock labelled solution and adjust if necessary to give 15 ± 2 x 10$^3$ cpm/50µL.

3) AFP Antiserum (Dakopotts, Copenhagen, Denmark):

   a) Stock solution: Dilute 50µL antiserum to 5mL with assay buffer (1 in 100 dilution). Dispense into 100µL aliquots and store at -20°C.
   b) Working solution: Thaw an aliquot of stock antiserum and dilute to 80mL with
assay buffer (final dilution 1 in 80000).

4) PEG (20% w/v):
   Dissolve 200g polyethylene glycol 6000 (Merck Ltd.) in about 800 mL distilled water. Make up to 1L and store at room temperature.

5) AFP-free serum:
   See section 3.1.1.1.1.

6) AFP Standards:
   a) Primary AFP standard (72/227) (NIBSC, Holly Hill, Hampstead, London): This freeze dried preparation contains approximately 2mL of pooled human cord serum designated to contain 98500 IU of AFP.
   b) Stock AFP standard: Dissolve the contents of the primary standard ampoule in 4.4mL assay buffer and leave in the dark at room temperature for 25 minutes. Dispense 1mL aliquots into 50mL volumetric flasks and freeze at -20°C.
   c) Working AFP standard: Thaw contents of a volumetric flask and dilute to 50mL with AFP-free serum to give 450 kU AFP/L. Into 10mL volumetric flasks, pipette 0.1, 0.2, 0.4, 1.0, 2.0, 3.5, and 5.0 mL of this standard and dilute to 10mL with AFP-free serum. Resulting standards are: 0, 4.5, 9, 18, 45, 90, 158, 225, and 450 kU AFP/L. Dispense these standards in 300µL aliquots into polystyrene tubes and store at -20°C. Thaw one tube of each standard for each AFP assay.

7) Quality Control sera:
   In-house; amniotic fluid diluted in AFP-free serum to give 3 levels. Store aliquots at -20°C.

Method:
1) Allow all reagents to attain room temperature.
2) To 50µL standard, serum or quality control (in duplicate), add 50µL of working AFP tracer.
3) To all tubes add 400µL of working antibody solution. Mix, cover and incubate overnight at room temperature.

4) To all tubes add 1mL of PEG. Mix thoroughly and incubate at room temperature for 25 minutes.

5) Centrifuge at 3000 rpm for 10 mins at room temperature.

6) Decant supernatant from all tubes except the total activity tube and leave inverted for 10 minutes.

7) Count all tubes for 1 minute.

8) Calculate results by B/B₀ vs LOG₁₀(concentration).

3.1.1.1.1 AFP-free Serum

Reagents:

1) Untreated powdered charcoal 250-350 mesh (Sigma Ltd.).

2) Molten Agar:

   Add 5g Oxoid No. 1 agar to 500mL distilled water and dissolve by steaming in a portable autoclave. Allow to cool to 60°C.

3) Acetone (Analar).

4) Defibrinated plasma.

Method:

1) To stirred cooled molten agar add 50g charcoal and mix thoroughly.

2) Add charcoal / agar to 1.5L stirred cold (0°C) acetone and mix.

3) Allow agar coated charcoal to settle and decant supernatant.

4) Wash (x2) with 1L fresh acetone at room temperature.

5) Separate charcoal by filtration using filter paper.

6) Spread the charcoal / agar and allow acetone to evaporate overnight.

7) Add the dried charcoal agar to 1L defibrinated plasma and mix
overnight at room temperature.
8) Filter and discard charcoal.
9) Centrifuge filtrate to remove fine particles of charcoal.
10) Add sodium azide to a final concentration of 0.1% (w/v)
11) Check AFP concentration of an aliquot.
12) Store at -20°C in appropriate aliquots.

3.1.1.2 Human Chorionic Gonadotropin assay (RGH-HCG)

**Principle:** In-house Immunoradiometric assay (IRMA).

**Reagents:**

1) Assay Buffer: 100µmol/L phosphate buffer containing 0.2% (w/v)
bovine serum albumin and 1% foetal calf serum.

Dissolve 32g anhydrous disodium hydrogen orthophosphate (Analar), 11g
sodium dihydrogen orthophosphate dihydrate (Analar), 10g bovine serum
albumin (Armour), and 10g sodium azide in 4500mL distilled water. Add 50 mL
fetal calf serum and adjust pH to 7.4. Make up to 5L with distilled water. Store at
4°C. Reagents from Merck Ltd, Poole, Dorset.

2) Tracer: ^{125}I-anti-β-HCG mouse monoclonal antibody (DPL Ltd.).

Reconstitute with 10.5mL distilled water.

3) Solid phase: Magnetic particle bound sheep polyclonal anti-HCG
antibody (UHW).

Working solid phase: dilute 1mL stock solid phase 1 in approximately 10mL of
assay buffer. Centrifuge and discard supernatant. Resuspend in x mL assay
buffer (x varies depending on antibody batch, approx 15-20 mL).
4) Wash reagent:

Make 40ml assay buffer up to 2L with distilled water and add 2mL Triton.

5) HCG Standards:

a) Stock standard: Human HCG from NIBSC reconstituted in assay buffer to give a final concentration of 500,000 IU/L. Aliquots stored frozen at -20°C.
b) Working standard: Dilute stock standard 1 in 250 to give top standard of 2000 IU/L and double dilute to give standard range of 1000, 500, 250, 125, 62.5 and 31.3 IU/L. Dilutions are prepared in 10mL volumes using assay buffer and stored in 0.2mL aliquots at -20°C.

6) Quality Control sera (in-house):

Three prediluted controls with values approximately 60, 300 and 1300 IU/L. One control (QC4) to be diluted as per samples with value approximately 22 IU/L.

Method:

1) Dilute serum samples (and QC4) 1 in 100 (10µL serum + 990µL) in assay buffer.
2) To 25µL standard, serum or quality control (in duplicate), add 100µL tracer.
3) Mix the solid phase well and add 100µL to each tube with a repeater eppendorf pipette.
4) Mix well and incubate for 30 mins at room temperature.
5) Add 1mL wash reagent to all tubes.
6) Mix well and place on magnets for 10 minutes.
7) Decant and repeat stages 5 and 6.
8) Decant and allow to drain for 10 mins.
9) Count all tubes for 1 minute.
10) Calculate results and express in IU/mL.
3.1.2 University Hospital of Wales Laboratory Methods

The laboratory at the University Hospital of Wales (UHW) measures AFP and total HCG concentrations using in-house assays and provides magnetically bound solid phase for use in the RGH-HCG assay. In later chapters the assays performed by the UHW will be referred to as UHW-AFP and UHW-HCG respectively.

3.1.2.1 Magnetic Solid Phase Preparation
(supplied to Royal Gwent for HCG assay)

Reagents:

Phosphate buffer (10μmol/L):

Dissolve 3.1g sodium dihydrogen orthophosphate dihydrate in 2L distilled water and adjust pH to 7.4.

Acetate buffer (10μmol/L):

Dissolve 2.7g sodium acetate and 17.5g sodium chloride in 2L distilled water and adjust pH to 5.5.

Method:

To 20mL of antiserum (AFP or HCG) add 20mL of phosphate buffer and 7.6g sodium sulphate. Mix for 2 hours and then centrifuge for 30 mins at 3000 rpm. Discard supernatant and dissolve deposit in 20mL distilled water. Dialyse this against acetate buffer for 4 days, changing buffer daily.

Wash 35mL of paramagnetic particles 4 times with methanol, then 4 times with acetate buffer. Add 8mL glutaraldehyde and 8mL acetate buffer and mix for 3 hours. Wash 4 times with acetate buffer. Add the
20mL of antiserum to the magnetic particles and mix overnight. Wash 4 times with 1% BSA, 4 times with phosphate buffer and dilute to 100mL with 1% BSA. Test in current assay before use to derive dilution factor required.

3.1.2.2 Serum α-fetoprotein assay (UHW-AFP)

**Principle:** In-house IRMA.

**Reagents:**

1) **Tracer:** $^{125}$I-anti-AFP polyclonal antibody (IDS Ltd.).

2) **Solid phase:** Magnetic particle bound sheep polyclonal anti-AFP antibody.

   Working solid phase: dilute 1mL stock solid phase 1 in approximately 10mL of assay buffer. Centrifuge and discard supernatant. Resuspend in x mL assay buffer (x varies depending on antibody batch, approx 15-20 mL).

3) **AFP Standards (IDS Ltd.):**

   Standards are calibrated against primary standard (72/227) and made up to give nominal values of 0, 6, 12, 24, 72, 144, and 288 kU/L.

4) **Quality control sera:**

   QC is Lyphochek (Biorad Ltd.,) which provides 3 levels (30, 100, 150 kU/L).

**Method:**

1) To 25μL standard (triplicate), serum or quality control (duplicate), add 100μL of tracer and 100μL of working magnetic solid phase.

2) Mix well and incubate overnight at room temperature.
3) Separate on a magnet. Discard supernatant and count magnetic particle bound radioactivity.

3.1.2.3 Human Chorionic Gonadotropin assay (UHW-HCG)

This assay is the same as the RGH-HCG assay with the following exceptions:

1) Sample dilution is 1 in 50 (20µL serum + 980µL buffer).

2) HCG Standards:
   Stock standard as RGH but 8 working standards are used having nominal values (0, 78, 156, 312, 625, 1250, 2500, 5000 IU/L).

3) Quality control sera:
   In-house: pooled pregnant serum at 3 levels treated as serum samples plus 3 Lyphochek control samples (Biorad Ltd.) which are not diluted.

3) Standards are assayed in triplicate.

3.1.3 Amersham international Assays

Amersham International (now called Kodak Diagnostics Ltd.) have developed assays which are optimised for the measurement of AFP, HCG and unconjugated oestriol in second trimester serum. All of these kits have catalogue numbers beginning with IM4 and therefore for simplicity these assays will be referred to as IM4-AFP, IM4-HCG, IM4-uE3. The initial work carried out on unconjugated oestriol was performed using a kit originally designed for use in third trimester pregnancies that
was modified to provide sensitivity at second trimester concentrations (Wald et al, 1988b: Canick et al, 1988). This assay had a catalogue number beginning with IM2 and will therefore be referred to as IM2-uE3.

3.1.3.1 Amerlex-M second trimester AFP RIA assay (IM4-AFP)

**Principle:** Radioimmunoassay (RIA).

**Reagents:**
1) Tracer: $^{125}$I-AFP solution.

2) Antiserum: Rabbit anti-AFP antiserum.

3) Separation system (second antibody): Magnetic particle bound anti-rabbit antiserum.

4) AFP standards: 7 vials to be reconstituted to 1mL with nominal concentrations: 0, 7.5, 15, 30, 70, 150, and 300 kU/L (Exact concentrations for each batch supplied with kit).

**Method:**
1) Allow specimens and reagents to reach room temperature before performing the assay.

2) To 100$\mu$L of standard, serum or quality control (in duplicate) add 100$\mu$L tracer and 100$\mu$L antiserum.

3) Mix well and incubate for 4 hours at 37°C.

4) Add 1ml of second antibody to all tubes using a repeater Eppendorf.

5) Incubate at room temperature for exactly 15 mins.
6) Separate on a magnet for 20 minutes.
7) Decant and drain for 10 minutes.
8) Count the radioactivity bound to the magnetic phase for sufficient time to accumulate at least 40,000 counts.

3.1.3.2 Amerlex-M second trimester HCG RIA assay (IM4-HCG)

**Principle**: Radioimmunoassay (RIA).

**Reagents**:
1) Tracer: ¹²⁵I-HCG solution.
2) Magnetic particle bound anti-HCG antiserum.
4) HCG standards: 7 vials to be reconstituted to 0.5mL with nominal concentrations: 0, 10, 20, 40, 80, 150, and 300 IU/mL (Exact concentrations for each batch supplied with kit).

**Method**:
1) Allow specimens and reagents to reach room temperature before performing the assay.
2) To 25µL of standard, serum or quality control (in duplicate) add 100µL tracer and 100µL antiserum.
3) Mix well and incubate for 1 hour at 37°C.
4) Separate on a magnet for 20 minutes.
5) Decant and drain for 10 minutes.
6) Count the radioactivity bound to the magnetic phase for sufficient time to accumulate at least 40,000 counts.
3.1.3.3 Amerlex-M second trimester uE3 RIA assay (IM4-uE3)

**Principle:** Radioimmunoassay (RIA).

**Reagents:**

1) Tracer: $^{125}$I-oestriol solution.

2) Magnetic particle bound anti-oestriol antiserum.

4) Oestriol standards: 7 vials to be reconstituted to 1mL with nominal concentrations: 0, 1, 2, 5, 10, 20, and 50 nmol/L (Exact concentrations for each batch supplied with kit).

**Method:**

1) Allow specimens and reagents to reach room temperature before performing the assay.

2) To 100µL of standard, serum or quality control (in duplicate) add 100µL tracer and 100µL antiserum.

3) Mix well and incubate for 1 hour at 37°C.

4) Separate on a magnet for 20 minutes.

5) Decant and drain for 10 minutes.

6) Count the radioactivity bound to the magnetic phase for sufficient time to accumulate at least 40,000 counts.
**3.1.3.4 Amerlex third trimester uE3 RIA assay (IM2-uE3)**

**Principle:** Radioimmunoassay (RIA).

**Reagents:**

1) Tracer: $^{125}$I-oestriol solution.

2) Anti-oestriol antiserum.

4) Oestriol standards: 6 vials to be reconstituted to 1mL with nominal concentrations: 0, 3, 6, 15, 30, and 50 nmol/L (Exact concentrations for each batch supplied with kit).

**Method:**

1) Allow specimens and reagents to reach room temperature before performing the assay.

2) To 20μL of standard, serum or quality control (in duplicate) add 100μL tracer and 500μL antiserum.

3) Mix well and incubate for 30 minutes at room temperature.

4) Separate on a magnet for 20 minutes.

5) Centrifuge at 1500g or greater for at least 15 minutes.

6) Decant the supernatant and allow to drain for at least 10 minutes.

7) Count the radioactivity of the pellet for sufficient time to accumulate at least 40,000 counts.
3.1.3.5 Amerlex third trimester uE3 RIA assay (Canick / Wald Modification) (IM2-uE3-Canick / Wald)

The Amerlex third trimester unconjugated oestriol kit does not have sufficient sensitivity to be used as provided by the manufacturer for assay of uE3 in second trimester samples. The kit was modified (Canick et al, 1988; Wald et al, 1988b) in order to demonstrate that the concentration of uE3 was lower in Down syndrome associated pregnancies. The modifications were as follows:

1) uE3 standards:
   The two highest standards were discarded and two new standards prepared by diluting the low standard (nominally 8 nmol/L) in zero standard. The resultant standards were: 0, 2, 4, 8, 21, 49 nmol/L (nominally).

2) Sample volume:
   This was increased from 20µL to 40µL.

3.1.4 CIS ELSA-free β-HCG assay (CIS-free β)

Principle: Two site IRMA.

Reagents:
1) ¹²⁵I-anti β-HCG monoclonal antibody.
2) Anti free β-HCG coated in excess onto a plastic insert (ELSA) in a tube.
3) Buffer.
4) Free β-HCG standards: 0, 2, 10, 20, and 50 ng/mL (nominal).
5) Control serum: (concentration lot dependent).
6) Wash reagent (Tween 20). Add 9mL to 3L distilled water.

**Methods:**

1) Add 200µL of buffer to all ELSA tubes.
2) Add 100µL standard, serum or quality control and Vortex mix.
3) Incubate for 1 hour at room temperature with constant agitation.
4) Wash (x3) with 3mL wash reagent. Allow to drain for 10 minutes after final wash.
5) Add 300µL tracer to each tube. Mix gently by Vortex mixer.
6) Incubate for 2 hours at room temperature with constant agitation.
7) Wash tubes (as step 4).
8) Measure radioactivity bound to solid phase.

### 3.2 MATHEMATICAL METHODS

#### 3.2.1 Age Related Risk Calculation

*(Cuckle et al, 1987)*

The risk of an individual woman having a pregnancy associated with Down syndrome depends to a great extent on her age. Cuckle *et al* performed a meta-analysis of 4528 affected and over 5 million unaffected pregnancies to derive an equation which could be used to calculate the age-related risk of Down syndrome for any individual woman.

**Equation 3.1**

\[
\text{Age Related Risk} = \frac{0.999373 - e^{(-16.2395 + 0.286 \times \text{Age})}}{0.000627 + e^{(-16.2395 + 0.286 \times \text{Age})}}
\]
The regression parameters in equation 3.1 were derived by logarithmic regression and have been critically re-examined (Whitehead, 1991) and compared with a pooled binomial regression method. Little difference was found between the 2 regression models (table 3.1) and therefore it was decided to retain the formula as originally described. A computer subroutine used to calculate Age risk is shown in Appendix 1, section 1.1.

3.2.2 Calculation of Multiples of the Median

The concentrations of the analytes used in screening are gestation age dependent, i.e. As gestation progresses, the concentration of each analyte changes (Fig 3.1). For example, the concentration of AFP increases and the concentration of HCG decreases with increasing gestational age. Therefore to enable comparison of results from women at different stages of pregnancy there are two possible methods. Either, a reference range can be derived for every stage of gestation or results can be standardised. The second approach is the most practical. The usual method for standardisation is calculation of the multiple of the median (MoM) which was originally described in 1977 (UK Collaborative Study of alpha-fetoprotein in relation to neural tube defects). This is performed by dividing a result by the median result that is found in women at the same stage of gestation. The MoM is actually a very poor method for standardisation because it does not satisfactorily remove inter-centre or inter-method biases (Macri et al, 1990f). The deficiencies of the MoM will be discussed in chapter 7.
Table 3.1

<table>
<thead>
<tr>
<th>Co-efficient</th>
<th>Cuckle et al</th>
<th>Whitehead</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.000627</td>
<td>0.000635</td>
</tr>
<tr>
<td>B</td>
<td>-16.2395</td>
<td>-16.3047</td>
</tr>
<tr>
<td>C</td>
<td>0.286</td>
<td>0.2875</td>
</tr>
</tbody>
</table>
Figure 3.1
Diagram to show change in median AFP, HCG and uE₃ concentrations with gestational age. uE₃ median values provided by Dr. P. Wood, Southampton General Hospital.
3.2.3 Weight correction of MoMs

The idea of weight correction of MoMs was proposed soon after the use of AFP for detection of neural tube defects became established. The underlying rationale is that AFP is water-soluble and is therefore distributed in the body water of the mother. Since at any given point in gestation, fetuses from different women tend to be of a similar size, their protein synthetic abilities also are likely to be similar. Thus, the total amount of AFP which can cross the placenta is likely to be similar irrespective of the size of the pregnant woman. The result of this is that heavy women tend to have low serum AFP concentrations and light women tend to have high serum AFP concentrations (Haddow et al, 1981: Wald et al, 1981).

A number of mathematical algorithms for weight correction of AFP have been proposed: AFP concentration cut-offs determined by arbitrary weight classifications (Haddow et al, 1981); linear regression to derive an adjustment factor (Palomaki et al, 1985); least squares logarithmic regression to derive an adjustment factor (Wald et al, 1981: Johnson et al, 1990); and multiple regression analysis which is also said to applicable for gestational age correction in the same equation (Johnson et al, 1984). Weight correction is however believed to be counter-productive by some other workers (Macri et al, 1986). Weight correction will be considered in chapter 5. A subroutine used to calculate corrected MoMs by the least squares logarithmic method is shown in Appendix 1, section 1.5.
3.2.4 Risk Calculation Methods

3.2.4.1 Linear Calculation Algorithm

DiMaio et al (1987) used a method for risk screening for Down syndrome which involved modification of the maternal age-related risk factor by a linear relationship (Baumgarten et al, 1985a: Baumgarten, 1985b) such that:

\[
\text{Risk} = 1 \text{ in } \frac{R}{r}
\]

and

\[
r = A + B \left( \frac{s}{m} \right)
\]

where \(R\) = Maternal age-related risk
\(r\) = Risk of Down syndrome
\(s\) = serum AFP concentration (corrected as required)
\(m\) = normative median for AFP for Gestation age
\(A\) = -2.15
\(B\) = 2.48

This method was only applied to screening with AFP alone and was claimed to detect 25 - 33% of cases of Down syndrome with a false positive rate of 5%.

3.2.4.2 Discriminant Function Method

This method was described by Nørgaard-Pedersen et al (1990). Using results from 328 normal and 42 Down syndrome samples, the linear function of the LOG\(_{10}\) MoM values for AFP, HCG and uE\(_3\) which best discriminated between the normal and Down syndrome serum results was derived by discriminant analysis. The function found was:
The value derived from this function was combined with the age-related risk by Baye’s theorem (Tabor et al, 1987) or if age was transformed and combined with D thus,

Equation 3.4
\[ T = 0.00325 e^{0.19 \cdot age} \]
\[ V = D - T \]

then all cases for which V was less than -0.51 the risk of Down syndrome was greater than 1 in 400. One possible disadvantage of this approach is that the discriminant function method is data dependent and different sets of data may behave differently. This risk algorithm will be examined in chapter 7.

3.2.4.3 Likelihood Ratio Method

The likelihood ratio method for calculating specific risks was first described in February 1987 by Palomaki and Haddow. They proposed that the Gaussian frequency distribution could be used to determine the chance that a given result came from an unaffected or an affected pregnancy with modification of the age-related risk by the likelihood ratio. Figure 3.2 shows the diagram from their paper which explains the idea. Since the results from affected and unaffected pregnancies come from normally distributed populations, the chance that any one result
Figure 3.2
Plot of Gaussian distributions of AFP in Down syndrome affected and unaffected pregnancies (Palomaki and Haddow, 1987).

Mean LOG(AFP) unaffected = 0.0 ± 0.19
Mean LOG(AFP) affected = -0.14 ± 0.21
comes from a particular distribution can be estimated by examining the
distance from the central point of that distribution i.e. by calculating how
many standard deviations away from the mean. However, since different
distributions vary in the size of their standard deviations, Palomaki and
Haddow used the 'height' above baseline in their calculation. Thus, the
relative chance that the result comes from one distribution or the other
can be determined as the ratio of the heights of distribution 1 and
distribution 2. It is not feasible to derive likelihood ratios graphically for
each result and therefore the Gaussian distribution formulae can be used
to determine the heights. If the heights for the 2 distributions are defined
as \( f_{\text{Normal}} \) and \( f_{\text{Down}} \) then the final Down syndrome risk is calculated thus:

\[
\text{Equation 3.5}
\text{Down syndrome risk} = \text{Age Risk} \times \frac{f_{\text{Normal}}}{f_{\text{Down}}}
\]

3.2.5 Statistical and Matrix Functions

3.2.5.1 Gaussian Distribution Functions


The Gaussian frequency distribution functions are well known
statistical formulae. If there is only one variable the univariate formula is

\[
\text{Equation 3.6}
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}
\]

where:

\( x, y \) = parameter
\( \mu \) = population mean
\( \sigma \) = population Std Dev
\( \rho \) = correlation coefficient
required (Equation 3.6). When there are two variables the bivariate formula is used (Equation 3.7). If there are more than 2 variables, specific formulae becomes excessively long and the general formula must be used (Equation 3.8). This formula requires an understanding of matrix mathematics which will be described next.

\[
\begin{align*}
\text{Equation 3.7} \\
f(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2} \left( \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho \left( \frac{x-\mu_x}{\sigma_x} \right) \left( \frac{y-\mu_y}{\sigma_y} \right) \right)}
\end{align*}
\]

Computer subroutines used to calculate Gaussian 'heights' are shown in Appendix 1, sections A1.6, 1.7 and 1.8.

\[
\begin{align*}
\text{Equation 3.8} \\
f(x_1,x_2..x_n) &= (2\pi)^{n/2} \cdot |V^{-1}| \cdot e^{-0.5 (x - \mu)^T V^{-1} (x - \mu)^T} 
\end{align*}
\]

3.2.5.2 Matrix Mathematics

A matrix is a group of numbers in a specific format which acts with other matrices in a very specific way. A matrix is formed from a variable number of rows and columns. The simplest form of matrix only has one row and one column and acts just as numbers do in every day life (i.e. the arithmetic which we all know and use is actually a special case of matrix mathematics).
3.2.5.2.1 Matrix Notation and Orientation

Before matrices can be manipulated it is important to know the correct mathematical notation which is used to denote different types of matrix.

\( X \): This represents the variable \( X \) which is a matrix variable as denoted by underlining the variable name.

\( X^{-1} \): This is the inverse of the matrix \( X \). The inverse is that matrix which when multiplied with the original matrix gives the identity matrix. This will be explained in section 3.2.5.2.5.

\( |X| \): This represents the determinant of the matrix. This will be described in section 3.2.5.2.4.

\( X' \): This represents the matrix \( X \) after it has been transposed. This means that its rows and columns have been switched around i.e. a matrix with 3 rows and 1 column becomes a matrix with 1 row and 3 columns.

To enable operations to be performed with matrices, the involved matrices must be 'compatible'. Compatibility is assessed by comparing the 'order' of different matrices. Order is written as \( (a \times b) \) where \( a \) refers to the number of rows in the matrix and \( b \) refers to the number of columns. Thus a \((3 \times 1)\) matrix has 3 rows and 1 column, a \((3 \times 3)\) matrix has 3 rows and columns and a \((1 \times 3)\) matrix has 1 row and 3 columns. The importance of order will be explained where required.

3.2.5.2.2 Matrix Addition and Subtraction

Only matrices which are of the same order can be added together or
subtracted from each other. The process is then very similar to ordinary mathematics with subtraction and addition being performed identically. For example, equation 3.9 shows the result of addition of two (3 x 1) matrices:

\[
\begin{pmatrix}
    x \\
    y \\
    z
\end{pmatrix}
+ 
\begin{pmatrix}
    m \\
    n \\
    o
\end{pmatrix}
= 
\begin{pmatrix}
    x + m \\
    y + n \\
    z + o
\end{pmatrix}
\]

### 3.2.5.2.3 Matrix Multiplication

Matrix multiplication is a more complex operation which can only take place if the participating matrices are compatible. Compatibility and the size of the product matrix can be determined by examining the orders of the participating matrices. For example, a (1 x 3) matrix can be multiplied by a (3 x 3) matrix but this procedure cannot occur in the other direction. Comparison of orders is performed by comparing the number of columns in the first matrix with the number of rows in the second. If these are equal then the first matrix may be multiplied by the second. If they are not equal it is impossible to multiply the matrix pair. The product matrix will have the same number of rows as the first matrix and the same number of columns as the second matrix. Thus in the example above, a (1 x 3) matrix can be multiplied by a (3 x 3) matrix and will result in a product matrix having 1 row and 3 columns, i.e. a (1 x 3) matrix.

When two compatible matrices are multiplied it is easiest to
understand the process involved by imagining that the second matrix topples over and each column interacts with the matrix it lands on independently. Thus the first element in the product matrix is the sum of the products of each row element in the first matrix multiplied by the corresponding column element in the second matrix (Equation 3.10).

\[
\begin{bmatrix}
a & b & c \\
d & e & f \\
g & h & i \\
\end{bmatrix}
\begin{bmatrix}
a \\
d \\
g \\
\end{bmatrix}
= (Aa + Bb + Cc, Ad + Be + Cf, Ag + Bh + Ci)
\]

3.2.5.2.4 Determinants

The determinant is a special value which can be computed for square matrices (a square matrix is a matrix which has the same number of rows as columns). The formal definition of the determinant can be written as:

\[
\sum \ldots \sum (-1)^i \alpha_{i_1j_1} \alpha_{i_2j_2} \ldots \alpha_{i_nj_n}
\]

This can be otherwise stated as the sum of all of the possible products of \( m \) elements of the matrix \( A \) in which each row and each column contributes exactly one element, with each of these products being multiplied by \((-1)^i\) where \( i \) is the number of inversions of the subscripts.
on the elements from the 'natural' ordering in which row and column subscripts match. Fortunately calculation of determinants is simpler than this definition sounds. There are simple 'special case' methods for calculating determinants of \((2 \times 2)\) and \((3 \times 3)\) matrices which are:

\[
\text{Equation 3.12}
\]

For the \((2 \times 2)\) matrix \(A:\)
\[
\begin{pmatrix}
  a & b \\
  c & d
\end{pmatrix}
\]

\(|A| = ad - bc\)

For the \((3 \times 3)\) matrix \(B:\)
\[
\begin{pmatrix}
  a & b & c \\
  d & e & f \\
  g & h & i
\end{pmatrix}
\]

\(|B| = aei + bfg + cdh - ceg - fha - ibd\)

For larger matrices it is more difficult to calculate determinants directly and a computer program which progressively cuts the matrix up into smaller manageable parts may be required. The method used for these larger matrices can be demonstrated using a \((3 \times 3)\) matrix as an example and is called 'expansion by cofactors'. Cofactors are also needed in the calculation of the inverse of a matrix. The cofactor of an element of a matrix is derived by taking the determinant of the remaining part of the matrix excluding the row and column in which that element is included. Thus, using the example of matrix \(B\) as in equation 3.12, the cofactor of element \(a\) is the determinant of the \((2 \times 2)\) matrix containing elements \(e, f, h\) and \(i\), i.e.

\[
\text{Cofactor}(a) = \begin{vmatrix}
  e & f \\
  h & i
\end{vmatrix}
\]

Similarly the cofactor of element \(b\) would be derived from elements \(d, g, f\)
and i. Equation 3.13 shows how this is done for matrix $B$. The same principle is applied to larger matrices with appropriate degradation of the matrix to allow determinants of large segments. It can be seen in equation 3.13 that the second element in the series is negativised. To determine which elements should be negativised, the sum of row and column number is utilised, i.e. the element $b$ comes from row 1, column 2 giving a value of 3. The element is therefore negativised because its position is odd. This is more clearly visualised if it is considered that each element is multiplied by $-1$ raised to the power of the sum of row and column ($-1^3 = -1$). A subroutine which calculates determinants is shown in Appendix 1, section 1.9.

**3.2.5.2.5 Matrix Inversion**

In the exponential term of equation 3.8, the $V$ matrix has the superscript $-1$ which means that the inverse of the $V$ matrix is used. The inverse of a matrix is that matrix containing a set of values which when multiplied by the original matrix gives a special matrix called the identity matrix ($I$). The identity matrix is also a special type of matrix and if a matrix is multiplied by it, the result will be the original matrix. As previously explained, the simplest form of matrix is a $(1 \times 1)$ matrix. Thus, as an example, if the $(1 \times 1)$ matrix $A = (2)$, then the identity matrix ($I$) corresponding to this matrix $I = (1)$, such that $(1) \times (2) = (2)$. Similarly,
the inverse \((A^{-1})\) of matrix \(A = (0.5)\) because \((0.5) \times (2) = 1\) (i.e. the identity matrix). The inverse of a \((1 \times 1)\) matrix is simple to derive because it is 1 divided by the value of the matrix. The inverse of a \((2 \times 2)\) matrix is also easy to derive (equation 3.14).

\[
\text{Equation 3.14}
\]

\[
\text{If } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ then } A^{-1} = \frac{1}{|A|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}
\]

There is no special case applicable to inversion of larger matrices, so a complex series of operations must be performed. The first step is 'transposition' where elements from row a, column b are swapped with corresponding elements in row b, column a. Next, each element is replaced by its 'minor' equivalent (this is the same value as the cofactor described in section 3.2.5.2.4 on determinants but without negativisation). The minor values are then converted to cofactors by the negativisation process already described and multiplied by the inverse of the determinant of the original matrix. The entire process is shown in equation 3.15.

The calculation of the inverse of a matrix is vital to the likelihood ratio method if more than 2 screening parameters are used. Fortunately, the values that are used to construct the \(V\) matrix which has to be inverted are all constants, i.e. the \(V\) matrix is the covariance matrix and is derived from the population standard deviations \((\sigma_1...\sigma_n)\) and the correlation coefficients between parameters \((\rho_{x1x2}, \rho_{x1x3}, \rho_{x2x3})\). This means that the inverse of this matrix also consists of constant values and allows
the computer program which calculates trivariate Gaussian heights to be simplified. The covariance matrix, its corresponding cofactor matrix (coV) and the determinant of the covariance matrix are shown as equation 3.16. The inverse of the covariance matrix is found by dividing each element in coV by |V| but since these elements can be calculated numerically, it is unnecessary to show the full algebraic expression for V⁻¹. The values for each element in the covariance matrix and the determinant and inverse matrix for each covariance matrix examined will be stated where appropriate. Subroutines for derivation of inverse matrices are shown in appendix 1, section A1.10.
3.3 SIMULATION METHODS

3.3.1 Random Normal Deviate Method

Simulation of any event which has a known Gaussian distribution can be achieved by setting a target mean and applying random variation to it, such that the frequency of any particular variation is determined by a Gaussian frequency distribution having a mean of 0 and a standard deviation of 1 (described as $N(0,1)$). The variation that is applied is determined by multiplying the randomly determined Gaussian distribution value by the target standard deviation. Thus, if a target mean ($\mu$) and a target sd ($\sigma$) is used, the result will be a Gaussian distribution with mean $\mu$ and sd $\sigma$ ($N(\mu,\sigma)$).

Simulations may be performed either by using a random number generator which produces a square distribution between 0 and 1 and applying this to an equation that calculates $Z$ (the number of standard
deviations away from the mean) (equation 3.17) or by using two independent squarely distributed random numbers and processing these by the Polar Marsaglia-Bray method (Donovan, 1992). I used this latter method in all simulations using a PASCAL subroutine modified from the

**Equation 3.17**

To calculate $Z(p)$ (to an error $\pm 10^{-3}$), where $p$ is a probability in the upper tail of a Gaussian distribution.

Let $q = 1 - p$; then let $t = \sqrt{\log_e(1/q^2)}$

\[
Z(p) = t - \left( \frac{(a_0 + a_1t)}{(1 + b_1t + b_2t^2)} \right)
\]

where $a_0 = 2.30753$; $a_1 = 0.27061$; $b_1 = 0.99229$; $b_2 = 0.04481$

(Abramowitz and Stegun, 1964)

 originals supplied by Donovan (appendix 1, section A1.12). A second important equation allows probability to be estimated from a standard normal deviate (equation 3.18). This equation is used for performing the Lilliefors test in a spreadsheet (see section 3.4.4).

**Equation 3.18**

To calculate $p(x)$ (to an error $\pm 10^{-7}$), where $x$ is a standard normal deviate ($\infty \geq x \geq 0$).

\[
p = 1 - 0.5(1 + d_1x + d_2x^2 + d_3x^3 + d_4x^4 + d_5x^5 + d_6x^6)^{16}
\]

where $d_1 = 0.0498673470$; $d_2 = 0.0211410061$; $d_3 = 0.0032776263$; $d_4 = 0.0000380036$; $d_5 = 0.0000488906$; $d_6 = 0.0000053830$.

(Abramowitz and Stegun, 1964)
3.3.2 Cholesky Root Method for Correlated Variables

In Down screening there are several variables which have known correlations between them. Any simulation of Down syndrome data must therefore produce correlated data. This can be achieved by multiplying a matrix containing a number of raw random normal deviates by the Cholesky square root of the corresponding covariance matrix (Donovan, 1992). For example, if three screening variables are to be simulated, then three standard normal deviates are placed into a $(3 \times 1)$ matrix, this matrix is multiplied by the Cholesky root matrix which, being derived from the covariance matrix, converts the random normal deviate matrix into a matrix containing three correlated standard normal deviates appropriate to the parameters being simulated. These values can then be modified to give the final set of simulated values. Subroutines to perform this calculation, modified from the originals supplied by Donovan, are shown in appendix 1, sections A1.13 and 1.14.

3.4 Statistical Methods

Most of the statistical methods used in this thesis are standard techniques which are described in many statistical textbooks. The texts used during this study were: (Meyer: Morrison, 1970; Harris, 1975; Martin et al, 1975; Swinscow, 1978; Barnett, 1979; Conover, 1980; Sprent, 1989). However, a number of the tests I used are unusual and these will be described briefly here.
3.4.1 Deming's Regression

Determination of a relationship between two variables is usually carried out by least squares regression analysis. This can be used to determine the correlation coefficient which gives an idea of how closely related the two variables are and regression coefficients which can be used to determine the value of y for a given x or vice-versa. However, least squares regression has the basic assumption that one of the pair of variables is absolutely correct and error free. This cannot be the case in biological systems when there may be error in both axes and this means that least squares regression frequently gives the wrong regression coefficients (Wakkers et al, 1975: Cornbleet and Gochman, 1979: Brooks et al, 1982). When comparing two sets of biological measurements where there may be error in both axes, the correct procedure to use is Deming's regression which allows variance affecting both axes to be taken into consideration. The differences between Deming's and least squares regression are shown as equation 3.19. Wherever appropriate it will be stated whether least squares or Deming's regression has been used.

Equation 3.19 (Wakkers et al, 1975)

Special symbols used in Deming’s regression:

\[ Q_x = \sum (x_i - \bar{x})^2; \quad Q_y = \sum (y_i - \bar{y})^2; \quad Q_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) \]

\[ \bar{x} = \frac{\sum x_i}{n}; \quad \bar{y} = \frac{\sum y_i}{n} \]

and the variance ratio \( \lambda = \frac{(\text{Standard Deviation of Method 1})^2}{(\text{Standard Deviation of Method 2})^2} \)
### 3.4.2 Skewness and Kurtosis

These two factors measure how close a distribution is to being perfectly Gaussian (Sachs, 1984; Linnet, 1987; Lacher, 1989). The coefficient of skewness measures whether the modal point of the distribution is deviated to lie above or below the mean and the coefficient of kurtosis measures whether the distribution is taller and narrower or flatter and wider than a perfect Gaussian distribution. The ideal value for

---

**Equation 3.19 cont’d.**

Comparison between least squares regression and Deming’s regression.

**Least Squares Regression:**

<table>
<thead>
<tr>
<th></th>
<th>X on Y</th>
<th>Y on X</th>
</tr>
</thead>
<tbody>
<tr>
<td>b (intercept)</td>
<td>$Q_{xy}$</td>
<td>$Q_y$</td>
</tr>
<tr>
<td></td>
<td>$Q_x$</td>
<td>$Q_{xy}$</td>
</tr>
<tr>
<td>a (slope)</td>
<td>$y - Bx$</td>
<td>$y - Bx$</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0</td>
<td>$\sqrt{((Q_x - Q_{xy}^2/Q_y)/(n-2))}$</td>
</tr>
<tr>
<td>Method 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$\sqrt{((Q_y - Q_{xy}^2/Q_x)/(n-2))}$</td>
<td>0</td>
</tr>
<tr>
<td>Method 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Deming’s Regression:**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>b (intercept)</td>
<td>$(\lambda Q_y - Q_x + \sqrt{((Q_x - \lambda Q_y)^2 + 4\lambda Q_{xy}^2})}/2\lambda Q_{xy}$</td>
</tr>
<tr>
<td>a (slope)</td>
<td>$y - Bx$</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$\sqrt{[\lambda/(1 + \lambda b^2). ((Q_y - 2bQ_{xy} + b^2Q_x)/(n - 2))]}$</td>
</tr>
<tr>
<td>Method 1</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$\sqrt{[1/(1 + \lambda b^2). ((Q_y - 2bQ_{xy} + b^2Q_x)/(n - 2))]}$</td>
</tr>
<tr>
<td>Method 2</td>
<td></td>
</tr>
</tbody>
</table>
the coefficient of skewness is 0; the ideal value for the coefficient of kurtosis is 3. The methods for calculation of the coefficients of skewness and kurtosis are shown in equation 3.20.

\[
\text{Coefficient of skewness} = \frac{\sum (x - \bar{x})^3}{n} \sigma^3
\]

\[
\text{Coefficient of kurtosis} = \frac{\sum (x - \bar{x})^4}{n} - 3 \sigma^4
\]

### 3.4.3 Confidence Intervals for the Median

Since the Down screening algorithm uses medians in many areas, it is necessary to have some idea of the confidence limits for these medians. For small sample sizes \((n \leq 20)\), confidence limits are assessed by using binomial frequency distribution tables (Conover, 1980; Campbell and Gardner, 1988) by finding the exact probabilities corresponding to the upper and lower limits required. For larger sample sizes, such as would be used to determine medians for Down syndrome screening, the central limit theorem applies and an approximation can be used to derive confidence limits (Equation 3.21).

\[
\text{Lower Conf. Limit} = \frac{n - (N_1 - \alpha/2 \cdot \sqrt{n}/2)}{2}
\]

\[
\text{Upper Conf. Limit} = 1 + \frac{n + (N_1 - \alpha/2 \cdot \sqrt{n}/2)}{2}
\]

where \(N_1 - \alpha/2\) = the appropriate value for the \(100(1 - \alpha/2)\) percentile derived from normal distribution tables.
3.4.4 Kolgomorov-Smirnov type tests

These tests are used to compare non-parametrically the goodness of fit of two separate samples, i.e. to determine whether they could have come from the same population distribution. It is not necessary for the two samples to be population samples, one of the ‘samples’ may be theoretical, e.g. it is possible to compare a sample with the Gaussian distribution curve to determine whether the sample could have come from a Gaussian distribution having a mean and standard deviation similar to that of the sample under test. All of the Kolgomorov-Smirnov type tests have the same basic method: the sample value, either as its absolute value (Smirnov test (to determine whether two independent samples could have come from the same distribution)) or scaled as standard deviations (Lilliefors test (to determine whether a sample conforms to a Gaussian distribution)) is plotted against its cumulative frequency (on a scale of 0 to 1). The test statistic is calculated as the maximum difference in height between the two cumulative probability curves. The method for calculating critical values for the Lilliefors and Smirnov tests are shown in equation 3.22 (Conover, 1980: Sprent, 1989).

**Equation 3.22**

Critical values for Smirnov test (large sample approximation: 2-sided test) where \( m \) and \( n \) are sample sizes of samples 1 and 2 respectively

10% level \( \omega_{0.90} = 1.22 \sqrt{\frac{(m + n)}{mn}} \)
5% level \( \omega_{0.95} = 1.36 \sqrt{\frac{(m + n)}{mn}} \)
1% level \( \omega_{0.99} = 1.63 \sqrt{\frac{(m + n)}{mn}} \)

Critical values for Lilliefors test (\( n > 30 \))

10% level \( \omega_{0.05} = 0.805 / \sqrt{n} \)
5% level \( \omega_{0.95} = 0.886 / \sqrt{n} \)
1% level \( \omega_{0.99} = 1.031 / \sqrt{n} \)
3.5 Ultrasound Formulae

There have been numerous reports on conversion of various fetal measurements to gestation ages. I have used equations for conversion published by the British Medical Ultrasound Society, Fetal Measurements Working Party in October 1990. The equations for Crown-Rump Length (CRL), Biparietal Diameter (BPD) and Femur length (FL) gestation dating and the original references are shown in equation 3.23.

<table>
<thead>
<tr>
<th>Equation 3.23</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRL (Robinson and Fleming, 1975):</td>
</tr>
</tbody>
</table>
| \[
| GA = \frac{[8.052 \sqrt{\text{CRL}} + 23.73]}{7}
| \]

<table>
<thead>
<tr>
<th>BPD (Hadlock et al, 1982):</th>
</tr>
</thead>
</table>
| \[
| GA = 6.8954 + 0.26345BPD + 0.000008771(BPD)^3
| \]

<table>
<thead>
<tr>
<th>FL (Warda et al, 1985):</th>
</tr>
</thead>
</table>
| \[
| GA = 2.35301 + 0.023185FL - 0.00007804(FL)^2
| \]

where GA is in weeks and ultrasound measurement in mm.
chapter 4:

**POPULATION NORMS AND MEDIANS**

### 4.1 POPULATION NORMAL RANGES

- 4.1.1 Source of data
- 4.1.2 Age distribution
- 4.1.3 Weight distribution
- 4.1.4 Height distribution
- 4.1.5 Biparietal diameter distribution
- 4.1.6 AFP distribution
- 4.1.7 HCG distribution
- 4.1.8 Distributions of other analytes

### 4.2 DERIVATION OF MEDIANS

- 4.2.1 Methods for derivation of medians
- 4.2.2 Weekly medians
- 4.2.3 Weekly weighted regressed medians
- 4.2.4 BMUS dated, regressed medians
- 4.2.5 Comparison of different medians

### 4.3 DISTRIBUTION OF MULTIPLES OF THE MEDIAN

- 4.3.1 MoM distributions for different medians
- 4.3.2 Effectiveness of correction for gestation age
- 4.3.3 Normality of MoM distributions
- 4.3.4 Correlation of AFP MoMs with HCG MoMs

### 4.4 CONCLUSIONS
4.1  POPULATION NORMAL RANGES

4.1.1 Source of Data

The data used to generate the distributions described in this chapter was collected during the periods 1st February 1990 to 14 July 1991 (data set 1) and 15 July 1991 to 30th June 1992 (data set 2) by the chemical pathology laboratory of the Royal Gwent Hospital and was made available by Dr. M. Penney. The data was collected primarily for use in the prospective Down syndrome screening program provided for women in Gwent. Requests for Down syndrome screening are made on a special request form designed by myself and Dr. Penney to ensure that all of the data required for the screening program are supplied correctly (Fig 4.1).

Data set 1 contained 5439 sets of data from uncomplicated singleton pregnancies but after incomplete sets were removed, there were 5137 cases with age at expected date of delivery, weight, BPD, date of scan, date of sample collection, AFP and HCG concentrations (assayed using the in-house methods described in chapter 3, sections 3.1.1.1 and 3.1.1.2). A further 55 cases were removed because the difference between date of scan and date of sample was greater than 3 days, leaving 5082 cases in all for derivation of medians.

Data set 2 contained 3334 sets of data from uncomplicated singleton pregnancies but after incomplete sets were removed, there were 3297 cases with age at expected date of delivery, height, weight, BPD, date of scan, date of sample collection, AFP and HCG concentrations.
<table>
<thead>
<tr>
<th>Surname</th>
<th>Patient Registration No.</th>
<th>Hospital</th>
<th>Lab No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mrs/Miss/Ms</td>
<td>Date of birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forenames</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>Consultant/GP</td>
<td>Ward</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LMP:</th>
<th>Date of Scan:</th>
<th>BPD (mm)</th>
<th>Femur Length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Minimum data set required BPD and/or LMP)

<table>
<thead>
<tr>
<th>Maternal Ethnic Origin</th>
<th>History of Diabetes Mellitus in patient</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Caucasian</td>
<td>Previous Down's associated pregnancy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 - Afro-Caribbean</td>
<td>Down's assoc'd preg'y in 1st degree relative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 - Asian</td>
<td>Previous NTD assoc'd preg'y in patient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 - Other</td>
<td>or 1st degree relative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tick correct option</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>CLINICAL DETAILS: Maternal:</th>
<th>height (cm); weight (kg at time of sampling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturity at booking:</td>
<td>by dates (wks); by size (wks) Multiple birth</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Blood sample collected:</td>
<td>at booking; after scan Other Relevant Information:</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Drs Signature</td>
<td>Date of collection / / RGH - CP</td>
</tr>
</tbody>
</table>

---

**Figure 4.1.**
Royal Gwent Hospital Down syndrome screening program request form.
4.1.2 Age Distribution

The age at expected date of delivery was determined by assuming a gestation period of 40 weeks and adding the estimated remaining gestation period (40 - gestation age) onto the maternal age at date of sample determined from the maternal date of birth. The distribution of maternal ages for data set 1 is shown in table 4.1 and figure 4.2. The mean (± sd) age at expected date of delivery was 27.0 (± 4.98) years, the median age was 27.3 years.

4.1.3 Weight Distribution

Due to the effect of weight on AFP medians in NTD screening (chapter 3; section 3.2.3), all request for Down screening are expected to include the weight at time of sampling on the request form. This is provided for most women except for a very small number who weigh in excess of 130 kg, when a note is made of gross obesity exceeding the capacity of the scales available to the ante-natal clinic. The distribution of the weights of 5136 pregnant women in Gwent (data set 1) is shown in table 4.2 and figure 4.3. The mean (± sd) weight at time of sampling was 64.88 (± 12.47) kg and the median was 62 kg.

4.1.3 Height Distribution

Since the effect of weight on AFP medians in NTD screening was well known, it was decided to investigate whether height would also affect MoMs. Therefore, all requests for Down screening after mid-July 1991 were expected to include the maternal height on the request form.
Table 4.1.
Age distribution of 5137 pregnant women in Gwent.

<table>
<thead>
<tr>
<th>Age</th>
<th>Number of women</th>
<th>Frequency (%)</th>
<th>Cumulative Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>1</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>16</td>
<td>28</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td>17</td>
<td>40</td>
<td>0.78</td>
<td>1.34</td>
</tr>
<tr>
<td>18</td>
<td>88</td>
<td>1.71</td>
<td>3.06</td>
</tr>
<tr>
<td>19</td>
<td>137</td>
<td>2.67</td>
<td>5.73</td>
</tr>
<tr>
<td>20</td>
<td>188</td>
<td>3.66</td>
<td>9.38</td>
</tr>
<tr>
<td>21</td>
<td>201</td>
<td>3.91</td>
<td>13.30</td>
</tr>
<tr>
<td>22</td>
<td>279</td>
<td>5.43</td>
<td>18.73</td>
</tr>
<tr>
<td>23</td>
<td>318</td>
<td>6.19</td>
<td>24.92</td>
</tr>
<tr>
<td>24</td>
<td>328</td>
<td>6.39</td>
<td>31.30</td>
</tr>
<tr>
<td>25</td>
<td>403</td>
<td>7.84</td>
<td>39.15</td>
</tr>
<tr>
<td>26</td>
<td>446</td>
<td>8.68</td>
<td>47.83</td>
</tr>
<tr>
<td>27</td>
<td>434</td>
<td>8.45</td>
<td>56.28</td>
</tr>
<tr>
<td>28</td>
<td>396</td>
<td>7.71</td>
<td>63.99</td>
</tr>
<tr>
<td>29</td>
<td>362</td>
<td>7.05</td>
<td>71.03</td>
</tr>
<tr>
<td>30</td>
<td>335</td>
<td>6.52</td>
<td>77.55</td>
</tr>
<tr>
<td>31</td>
<td>239</td>
<td>4.65</td>
<td>82.21</td>
</tr>
<tr>
<td>32</td>
<td>210</td>
<td>4.09</td>
<td>86.30</td>
</tr>
<tr>
<td>33</td>
<td>161</td>
<td>3.13</td>
<td>89.43</td>
</tr>
<tr>
<td>34</td>
<td>134</td>
<td>2.61</td>
<td>92.04</td>
</tr>
<tr>
<td>35</td>
<td>101</td>
<td>1.97</td>
<td>94.00</td>
</tr>
<tr>
<td>36</td>
<td>97</td>
<td>1.89</td>
<td>95.89</td>
</tr>
<tr>
<td>37</td>
<td>68</td>
<td>1.32</td>
<td>97.22</td>
</tr>
<tr>
<td>38</td>
<td>52</td>
<td>1.01</td>
<td>98.23</td>
</tr>
<tr>
<td>39</td>
<td>38</td>
<td>0.74</td>
<td>98.97</td>
</tr>
<tr>
<td>40</td>
<td>19</td>
<td>0.37</td>
<td>99.34</td>
</tr>
<tr>
<td>41</td>
<td>15</td>
<td>0.29</td>
<td>99.63</td>
</tr>
<tr>
<td>42</td>
<td>4</td>
<td>0.08</td>
<td>99.71</td>
</tr>
<tr>
<td>43</td>
<td>7</td>
<td>0.14</td>
<td>99.84</td>
</tr>
<tr>
<td>44</td>
<td>4</td>
<td>0.08</td>
<td>99.92</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>0.02</td>
<td>99.94</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
<td>0.02</td>
<td>99.96</td>
</tr>
<tr>
<td>47</td>
<td>2</td>
<td>0.04</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total 5137
Figure 4.2.
Age distribution of the pregnant population of Gwent.
Table 4.2.
Weight distribution of 5137 pregnant women in Gwent.

<table>
<thead>
<tr>
<th>Weight (kg)</th>
<th>Number of women</th>
<th>Frequency (%)</th>
<th>Cumulative Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 35</td>
<td>4</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>35 - 40</td>
<td>50</td>
<td>0.97</td>
<td>1.05</td>
</tr>
<tr>
<td>40 - 45</td>
<td>241</td>
<td>4.69</td>
<td>5.74</td>
</tr>
<tr>
<td>45 - 50</td>
<td>625</td>
<td>12.17</td>
<td>17.90</td>
</tr>
<tr>
<td>50 - 55</td>
<td>1014</td>
<td>19.74</td>
<td>37.65</td>
</tr>
<tr>
<td>55 - 60</td>
<td>1002</td>
<td>19.51</td>
<td>57.15</td>
</tr>
<tr>
<td>60 - 65</td>
<td>784</td>
<td>15.26</td>
<td>72.42</td>
</tr>
<tr>
<td>65 - 70</td>
<td>523</td>
<td>10.18</td>
<td>82.60</td>
</tr>
<tr>
<td>70 - 75</td>
<td>308</td>
<td>6.00</td>
<td>88.59</td>
</tr>
<tr>
<td>75 - 80</td>
<td>192</td>
<td>3.74</td>
<td>92.33</td>
</tr>
<tr>
<td>80 - 85</td>
<td>137</td>
<td>2.67</td>
<td>95.00</td>
</tr>
<tr>
<td>85 - 90</td>
<td>106</td>
<td>2.06</td>
<td>97.06</td>
</tr>
<tr>
<td>90 - 95</td>
<td>54</td>
<td>1.05</td>
<td>98.11</td>
</tr>
<tr>
<td>95 - 100</td>
<td>40</td>
<td>0.78</td>
<td>98.89</td>
</tr>
<tr>
<td>100 - 105</td>
<td>19</td>
<td>0.37</td>
<td>99.26</td>
</tr>
<tr>
<td>105 - 110</td>
<td>12</td>
<td>0.23</td>
<td>99.49</td>
</tr>
<tr>
<td>110 - 115</td>
<td>13</td>
<td>0.25</td>
<td>99.75</td>
</tr>
<tr>
<td>115 - 120</td>
<td>8</td>
<td>0.16</td>
<td>99.90</td>
</tr>
<tr>
<td>120 - 125</td>
<td>4</td>
<td>0.08</td>
<td>99.98</td>
</tr>
<tr>
<td>125 - 130</td>
<td>1</td>
<td>0.02</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Figure 4.3.
Weight distribution of the pregnant population of Gwent.
The distribution of the heights of 2921 pregnant women in Gwent (data set 2) is shown in table 4.3 and figure 4.4. The mean (± sd) height was 1.62 (± 0.06) m and the median was 1.63 m.

The correlation between weight and height was examined by least squares regression (Fig 4.5) and the regression coefficients were:

\[
\text{Slope} = 59.77; \quad \text{Intercept} = -30.93; \quad r = 0.2955; \quad t = 16.71; \quad P << 0.0001
\]

4.1.5 Biparietal Diameter Distribution

In the RGH ante-natal Down screening program, gestation dates are almost exclusively determined by BPD measurement within 3 days of blood sampling (5408 cases of 5439; 99.4%). Therefore, the distribution of LMP dates is not available. The method of ultrasound dating changed during mid-1991 and both methods of dating used will be examined. The original dating method utilised BPD size bands such that any BPD size within a band indicated a week of gestation; when this was changed, the British Medical Ultrasound Society recommended formula for conversion of BPD to a decimal week was adopted (Chapter 3: section 3.5; Equation 3.22). The distribution of BPD sizes with the corresponding weeks of gestation are shown in table 4.4 and figure 4.6. The mean (± sd) BPD was 38.37 (± 2.87) mm and the mean gestation date determined by weekly size band was 16.54 (± 0.84) weeks and by BMUS formula was 17.51 (± 0.87) weeks.
Table 4.3.
Height distribution of 2921 pregnant women in Gwent.

<table>
<thead>
<tr>
<th>Height (m)</th>
<th>Number of women</th>
<th>Frequency (%)</th>
<th>Cumulative Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.05 - 1.10</td>
<td>1</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>1.10 - 1.15</td>
<td>0</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>1.15 - 1.20</td>
<td>0</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>1.20 - 1.25</td>
<td>0</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>1.25 - 1.30</td>
<td>1</td>
<td>0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>1.30 - 1.35</td>
<td>2</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>1.35 - 1.40</td>
<td>3</td>
<td>0.10</td>
<td>0.24</td>
</tr>
<tr>
<td>1.40 - 1.45</td>
<td>8</td>
<td>0.27</td>
<td>0.51</td>
</tr>
<tr>
<td>1.45 - 1.50</td>
<td>76</td>
<td>2.60</td>
<td>3.12</td>
</tr>
<tr>
<td>1.50 - 1.55</td>
<td>342</td>
<td>11.71</td>
<td>14.82</td>
</tr>
<tr>
<td>1.55 - 1.60</td>
<td>822</td>
<td>28.14</td>
<td>42.96</td>
</tr>
<tr>
<td>1.60 - 1.65</td>
<td>885</td>
<td>30.30</td>
<td>73.26</td>
</tr>
<tr>
<td>1.65 - 1.70</td>
<td>560</td>
<td>19.17</td>
<td>92.43</td>
</tr>
<tr>
<td>1.70 - 1.75</td>
<td>175</td>
<td>5.99</td>
<td>98.43</td>
</tr>
<tr>
<td>1.75 - 1.80</td>
<td>39</td>
<td>1.34</td>
<td>99.76</td>
</tr>
<tr>
<td>1.80 - 1.85</td>
<td>6</td>
<td>0.21</td>
<td>99.97</td>
</tr>
<tr>
<td>1.85 - 1.90</td>
<td>1</td>
<td>0.03</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Figure 4.4.
Height distribution of the pregnant population of Gwent.
Figure 4.5.
Relationship between weight and height. The line is the regression line derived in section 4.1.4. Only 200 individual data points have been shown for clarity.
Table 4.4.
Distribution of biparietal diameters and relationship to gestation date.

<table>
<thead>
<tr>
<th>BPD (mm)</th>
<th>Number of women</th>
<th>Frequency (%)</th>
<th>Cumulative Frequency (%)</th>
<th>Date by Size band (weeks)</th>
<th>Date by BMUS formula (week + day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>2</td>
<td>0.04</td>
<td>0.04</td>
<td>15</td>
<td>15 + 0</td>
</tr>
<tr>
<td>31</td>
<td>28</td>
<td>0.55</td>
<td>0.58</td>
<td>15</td>
<td>15 + 2</td>
</tr>
<tr>
<td>32</td>
<td>49</td>
<td>0.95</td>
<td>1.54</td>
<td>15</td>
<td>15 + 4</td>
</tr>
<tr>
<td>33</td>
<td>138</td>
<td>2.69</td>
<td>4.22</td>
<td>15</td>
<td>15 + 6</td>
</tr>
<tr>
<td>34</td>
<td>180</td>
<td>3.50</td>
<td>7.73</td>
<td>15</td>
<td>16 + 1</td>
</tr>
<tr>
<td>35</td>
<td>324</td>
<td>6.31</td>
<td>14.04</td>
<td>16</td>
<td>16 + 3.5</td>
</tr>
<tr>
<td>36</td>
<td>742</td>
<td>14.44</td>
<td>28.48</td>
<td>16</td>
<td>16 + 6</td>
</tr>
<tr>
<td>37</td>
<td>743</td>
<td>14.46</td>
<td>42.94</td>
<td>16</td>
<td>17 + 1</td>
</tr>
<tr>
<td>38</td>
<td>726</td>
<td>14.13</td>
<td>57.08</td>
<td>16</td>
<td>17 + 3</td>
</tr>
<tr>
<td>39</td>
<td>525</td>
<td>10.22</td>
<td>67.30</td>
<td>17</td>
<td>17 + 5</td>
</tr>
<tr>
<td>40</td>
<td>616</td>
<td>11.99</td>
<td>79.29</td>
<td>17</td>
<td>18 + 0</td>
</tr>
<tr>
<td>41</td>
<td>385</td>
<td>7.49</td>
<td>86.78</td>
<td>17</td>
<td>18 + 2</td>
</tr>
<tr>
<td>42</td>
<td>274</td>
<td>5.33</td>
<td>92.12</td>
<td>18</td>
<td>18 + 4</td>
</tr>
<tr>
<td>43</td>
<td>186</td>
<td>3.62</td>
<td>95.73</td>
<td>18</td>
<td>18 + 6.5</td>
</tr>
<tr>
<td>44</td>
<td>90</td>
<td>1.75</td>
<td>97.49</td>
<td>18</td>
<td>19 + 2</td>
</tr>
<tr>
<td>45</td>
<td>56</td>
<td>1.09</td>
<td>98.58</td>
<td>18</td>
<td>19 + 4</td>
</tr>
<tr>
<td>46</td>
<td>34</td>
<td>0.66</td>
<td>99.24</td>
<td>19</td>
<td>19 + 6</td>
</tr>
<tr>
<td>47</td>
<td>25</td>
<td>0.49</td>
<td>99.73</td>
<td>19</td>
<td>20 + 1.5</td>
</tr>
<tr>
<td>48</td>
<td>11</td>
<td>0.21</td>
<td>99.94</td>
<td>19</td>
<td>20 + 4</td>
</tr>
<tr>
<td>49</td>
<td>3</td>
<td>0.06</td>
<td>100.00</td>
<td>19</td>
<td>20 + 6</td>
</tr>
</tbody>
</table>
Figure 4.6.
Foetal BPD distribution in 5137 pregnant women in Gwent.
4.1.6 AFP Distribution

The AFP concentrations measured in 5082 consecutive samples ranged from 2 - 999 kU/L with a mean (± sd) of 37.3 ± 22.95 kU/L. Since the distribution of uncorrected analytes is only of significance in relation to assay linearity, distribution data will only be presented numerically. For AFP, the distribution of concentrations is shown in table 4.5.

4.1.7 HCG Distribution

The HCG concentrations measured in 5082 consecutive samples ranged from 1 - 366 IU/L with a mean (± sd) of 22.49 ± 15.24 kU/L. For HCG, the distribution of concentrations is shown in table 4.6.

4.1.8 Distributions of other analytes

No other analytes were assayed routinely: uE₃ and free β HCG were assayed on selected samples for specific research purposes but since there were only a small number of these highly selected samples, the population distributions found were not representative and have not therefore been derived.

4.2 Derivation Of Medians

4.2.1 Methods for derivation of medians

Three methods were used to derive the medians required for calculation of MoMs for AFP and HCG. These were:

1) Gestation dating by week size band (table 4.4) with median
Table 4.5.
Distribution of AFP concentrations in 5082 consecutive samples.

<table>
<thead>
<tr>
<th>AFP kU/L</th>
<th>Number of Samples</th>
<th>Frequency (%)</th>
<th>Cumulative Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10</td>
<td>10</td>
<td>0.20</td>
<td>0.2</td>
</tr>
<tr>
<td>10 - 15</td>
<td>87</td>
<td>1.71</td>
<td>1.91</td>
</tr>
<tr>
<td>15 - 20</td>
<td>312</td>
<td>6.14</td>
<td>8.05</td>
</tr>
<tr>
<td>20 - 25</td>
<td>655</td>
<td>12.89</td>
<td>20.94</td>
</tr>
<tr>
<td>25 - 30</td>
<td>838</td>
<td>16.49</td>
<td>37.43</td>
</tr>
<tr>
<td>30 - 35</td>
<td>789</td>
<td>15.53</td>
<td>52.96</td>
</tr>
<tr>
<td>35 - 40</td>
<td>626</td>
<td>12.32</td>
<td>65.28</td>
</tr>
<tr>
<td>40 - 45</td>
<td>530</td>
<td>10.43</td>
<td>75.71</td>
</tr>
<tr>
<td>45 - 50</td>
<td>364</td>
<td>7.16</td>
<td>82.88</td>
</tr>
<tr>
<td>50 - 55</td>
<td>245</td>
<td>4.82</td>
<td>87.70</td>
</tr>
<tr>
<td>55 - 60</td>
<td>190</td>
<td>3.74</td>
<td>91.44</td>
</tr>
<tr>
<td>60 - 65</td>
<td>140</td>
<td>2.76</td>
<td>94.19</td>
</tr>
<tr>
<td>65 - 70</td>
<td>81</td>
<td>1.59</td>
<td>95.79</td>
</tr>
<tr>
<td>70 - 75</td>
<td>70</td>
<td>1.38</td>
<td>97.17</td>
</tr>
<tr>
<td>75 - 80</td>
<td>48</td>
<td>0.94</td>
<td>98.11</td>
</tr>
<tr>
<td>80 - 85</td>
<td>27</td>
<td>0.53</td>
<td>98.64</td>
</tr>
<tr>
<td>85 - 90</td>
<td>20</td>
<td>0.39</td>
<td>99.04</td>
</tr>
<tr>
<td>90 - 95</td>
<td>10</td>
<td>0.20</td>
<td>99.23</td>
</tr>
<tr>
<td>95 - 100</td>
<td>6</td>
<td>0.12</td>
<td>99.35</td>
</tr>
<tr>
<td>100 - 105</td>
<td>4</td>
<td>0.08</td>
<td>99.43</td>
</tr>
<tr>
<td>105 - 110</td>
<td>3</td>
<td>0.06</td>
<td>99.49</td>
</tr>
<tr>
<td>110 - 115</td>
<td>6</td>
<td>0.12</td>
<td>99.61</td>
</tr>
<tr>
<td>115 - 120</td>
<td>1</td>
<td>0.02</td>
<td>99.63</td>
</tr>
<tr>
<td>&gt;120</td>
<td>20</td>
<td>0.37</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Table 4.6.
Distribution of HCG concentrations in 5082 consecutive samples.

<table>
<thead>
<tr>
<th>HCG IU/L</th>
<th>Number of Samples</th>
<th>Frequency (%)</th>
<th>Cumulative Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 5</td>
<td>30</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>5 - 10</td>
<td>386</td>
<td>7.60</td>
<td>8.19</td>
</tr>
<tr>
<td>10 - 15</td>
<td>812</td>
<td>15.98</td>
<td>24.17</td>
</tr>
<tr>
<td>15 - 20</td>
<td>972</td>
<td>19.13</td>
<td>43.30</td>
</tr>
<tr>
<td>20 - 25</td>
<td>916</td>
<td>18.03</td>
<td>61.33</td>
</tr>
<tr>
<td>25 - 30</td>
<td>565</td>
<td>11.12</td>
<td>72.45</td>
</tr>
<tr>
<td>30 - 35</td>
<td>459</td>
<td>9.03</td>
<td>81.48</td>
</tr>
<tr>
<td>35 - 40</td>
<td>315</td>
<td>6.20</td>
<td>87.68</td>
</tr>
<tr>
<td>40 - 45</td>
<td>216</td>
<td>4.25</td>
<td>91.93</td>
</tr>
<tr>
<td>45 - 50</td>
<td>143</td>
<td>2.81</td>
<td>94.75</td>
</tr>
<tr>
<td>50 - 55</td>
<td>71</td>
<td>1.40</td>
<td>96.14</td>
</tr>
<tr>
<td>55 - 60</td>
<td>64</td>
<td>1.26</td>
<td>97.40</td>
</tr>
<tr>
<td>60 - 65</td>
<td>20</td>
<td>0.39</td>
<td>97.80</td>
</tr>
<tr>
<td>65 - 70</td>
<td>34</td>
<td>0.67</td>
<td>98.46</td>
</tr>
<tr>
<td>70 - 75</td>
<td>15</td>
<td>0.29</td>
<td>98.76</td>
</tr>
<tr>
<td>75 - 80</td>
<td>21</td>
<td>0.41</td>
<td>99.17</td>
</tr>
<tr>
<td>80 - 85</td>
<td>7</td>
<td>0.14</td>
<td>99.31</td>
</tr>
<tr>
<td>85 - 90</td>
<td>10</td>
<td>0.20</td>
<td>99.51</td>
</tr>
<tr>
<td>90 - 95</td>
<td>1</td>
<td>0.02</td>
<td>99.53</td>
</tr>
<tr>
<td>95 - 100</td>
<td>5</td>
<td>0.10</td>
<td>99.63</td>
</tr>
<tr>
<td>&gt;100</td>
<td>20</td>
<td>0.37</td>
<td>100.00</td>
</tr>
</tbody>
</table>
derived by sorting all results in a size band into order and examining the result at the 50th percentile.

2) Weighted regression of gestation week derived from size band against median derived above or \( \log_{10} \) of median derived above.

3) Regression of BMUS formula derived gestation date against AFP of HCG concentration or \( \log_{10} \) of AFP or HCG concentration.

4.2.2 Weekly medians

Table 4.7 shows the AFP and HCG concentrations at each week at the 90th, 75th, 50th, 25th and 10th percentiles. The value at the 50th percentile is the median. Table 4.8 contains statistical information about the distribution of AFP and HCG results at each week.

4.2.3 Weekly Weighted Regressed Medians

Table 4.9 shows the least squares regression parameters for gestation age determined by BPD size bands versus AFP or HCG median and \( \log_{10} \) of AFP or HCG median.

4.2.4 BMUS dated, Regressed Medians

Table 4.10 shows the least squares regression parameters for gestation age determined by the BMUS formula relating BPD to dates versus AFP or HCG concentration or \( \log_{10} \) of AFP or HCG concentration.
Table 4.7. Percentile distribution of AFP and HCG concentrations at different weeks of gestation (determined by weekly size band). \( n \) is the number of results in each week band and is the same for both AFP and HCG.

### AFP (kU/L)

<table>
<thead>
<tr>
<th>Week of Gestation</th>
<th>Percentile</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>90</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>17</td>
<td>22</td>
<td>29</td>
<td>39</td>
<td>51</td>
<td>329</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>24</td>
<td>32</td>
<td>41</td>
<td>53</td>
<td>2410</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>22</td>
<td>27</td>
<td>34</td>
<td>46</td>
<td>59</td>
<td>1656</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>24</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>66</td>
<td>610</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>29</td>
<td>38</td>
<td>48</td>
<td>62</td>
<td>72</td>
<td>77</td>
<td></td>
</tr>
</tbody>
</table>

### HCG (IU/L)

<table>
<thead>
<tr>
<th>Week of Gestation</th>
<th>Percentile</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>13</td>
<td>19</td>
<td>26</td>
<td>38</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>16</td>
<td>22</td>
<td>33</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>14</td>
<td>20</td>
<td>29</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>12</td>
<td>18</td>
<td>24</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>9</td>
<td>15</td>
<td>23</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.8. Statistical details of the distributions of AFP and HCG at each week of gestation.

**AFP (kU/L)**

<table>
<thead>
<tr>
<th>Week of Gestation</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness coefficient</th>
<th>Kurtosis coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>29</td>
<td>32.90</td>
<td>19.14</td>
<td>4.10</td>
<td>32.60</td>
</tr>
<tr>
<td>16</td>
<td>32</td>
<td>34.76</td>
<td>15.76</td>
<td>2.69</td>
<td>20.57</td>
</tr>
<tr>
<td>17</td>
<td>34</td>
<td>37.54</td>
<td>15.17</td>
<td>1.25</td>
<td>6.63</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
<td>42.73</td>
<td>18.54</td>
<td>1.77</td>
<td>8.66</td>
</tr>
<tr>
<td>19</td>
<td>48</td>
<td>49.70</td>
<td>16.80</td>
<td>0.36</td>
<td>2.57</td>
</tr>
<tr>
<td>Overall</td>
<td>37.23</td>
<td>22.95</td>
<td>17.93</td>
<td>667.30</td>
<td></td>
</tr>
</tbody>
</table>

**HCG (IU/L)**

<table>
<thead>
<tr>
<th>Week of Gestation</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness coefficient</th>
<th>Kurtosis coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>26</td>
<td>30.12</td>
<td>16.72</td>
<td>1.86</td>
<td>8.50</td>
</tr>
<tr>
<td>16</td>
<td>22</td>
<td>26.09</td>
<td>16.72</td>
<td>5.54</td>
<td>87.86</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>22.47</td>
<td>12.17</td>
<td>1.48</td>
<td>6.46</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>19.74</td>
<td>10.74</td>
<td>1.31</td>
<td>5.30</td>
</tr>
<tr>
<td>19</td>
<td>15</td>
<td>18.62</td>
<td>15.61</td>
<td>4.56</td>
<td>30.61</td>
</tr>
<tr>
<td>Overall</td>
<td>24.49</td>
<td>15.23</td>
<td>4.20</td>
<td>64.49</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.9.
Regression parameters for least squares weighted regression of gestation week versus median.

<table>
<thead>
<tr>
<th></th>
<th>Slope</th>
<th>Intercept</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>3.6514</td>
<td>-26.76</td>
<td>0.93985</td>
</tr>
<tr>
<td>LOG\textsubscript{10}(AFP)</td>
<td>0.044646</td>
<td>0.78652</td>
<td>0.9585</td>
</tr>
<tr>
<td>HCG</td>
<td>-2.33522</td>
<td>59.66103</td>
<td>-0.9773</td>
</tr>
<tr>
<td>LOG\textsubscript{10}(HCG)</td>
<td>-0.049085</td>
<td>2.13284</td>
<td>-0.98583</td>
</tr>
</tbody>
</table>
Table 4.10.
Regression parameters for least squares regression of BMUS formula derived gestation date versus concentration. In all cases t indicates that the regression is significant with P << 0.0001.

<table>
<thead>
<tr>
<th></th>
<th>Slope</th>
<th>Intercept</th>
<th>r</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>3.9026</td>
<td>-31.1</td>
<td>0.1482</td>
<td>10.68</td>
</tr>
<tr>
<td>LOG₁₀(AFP)</td>
<td>0.04600</td>
<td>0.7247</td>
<td>0.22111</td>
<td>16.16</td>
</tr>
<tr>
<td>HCG</td>
<td>-3.13503</td>
<td>79.3872</td>
<td>-0.17937</td>
<td>12.99</td>
</tr>
<tr>
<td>LOG₁₀(HCG)</td>
<td>-0.0586372</td>
<td>2.3485</td>
<td>-0.20952</td>
<td>15.27</td>
</tr>
</tbody>
</table>
4.2.5 Comparison of different medians

Table 4.11 shows medians derived using each of the different methods for weeks 15 - 19. There are obvious differences between the sets of medians which may be accentuated by the differences between week size band and BMUS formula derived gestation dates. To compensate for this, the medians can be plotted against the BPD corresponding to each date (Figures 4.7 and 4.8 show this for AFP and HCG respectively). In both cases, the graphs show that medians derived by regression of analyte concentration against BMUS formula derived gestation age (BMUS-### medians) are significantly different to medians derived by the other methods and this suggests that this method should not be used. This can be further examined by deriving multiples of the median using each set of medians.

4.3 Distribution of multiples of the median

4.3.1 MoM Distribution for Different Derivations of Median

Tables 4.12 and 4.13 show the distributions of MoMs for the 5082 serum results of data set 1 for AFP and HCG respectively. Evaluation of the figures for AFP show that the distribution of MoMs derived using BMUS-AFP medians is significantly different from the MoM distributions derived from the other 4 derivations of medians, which are all remarkably similar (approximate ranges 0.354 - 2.381 versus 0.385 - 2.62).

Furthermore, the point at which the AFP MoM = 1 is approximately at the 50th percentile for four of the median derivation methods and is at the 60th percentile for the other (the BMUS-AFP median).
Table 4.11.
Medians for AFP and HCG for gestation weeks 15 - 19, derived by
weekly size band, weighted regression of weekly median and regression
of BMUS formula gestation date vs concentration.

<table>
<thead>
<tr>
<th>AFP (kU/L)</th>
<th>Gestation Week</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Weekly</td>
<td>29</td>
</tr>
<tr>
<td>Weighted (median)</td>
<td>28.0</td>
</tr>
<tr>
<td>Weighted LOG$_{10}$(median)</td>
<td>28.6</td>
</tr>
<tr>
<td>BMUS date vs concentration</td>
<td>27.4</td>
</tr>
<tr>
<td>BMUS date vs LOG$_{10}$(conc$^a$)</td>
<td>26.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HCG (IU/L)</th>
<th>Gestation Week</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Weekly</td>
<td>26</td>
</tr>
<tr>
<td>Weighted (median)</td>
<td>24.6</td>
</tr>
<tr>
<td>Weighted LOG$_{10}$(median)</td>
<td>24.9</td>
</tr>
<tr>
<td>BMUS date vs concentration</td>
<td>32.4</td>
</tr>
<tr>
<td>BMUS date vs LOG$_{10}$(conc$^a$)</td>
<td>29.4</td>
</tr>
</tbody>
</table>
Figure 4.7.
Graphs to show differences in AFP medians derived by different methods. The stepped line shows the weekly median values.
Figure 4.8.
Graphs to show differences in HCG medians derived by different methods. The stepped line shows the weekly median values.
Table 4.12.
MoM distribution for 5082 consecutive AFP results. Distributions shown are MoMs derived using: weekly medians; weighted regressed weekly medians; weighted regressed LOG(weekly medians); regression of AFP concentration against BMUS formula derived gestation age; regression of LOG$_{10}$(AFP) concentration against BMUS formula derived gestation age.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Weekly median</th>
<th>Weighted regressed median</th>
<th>Weighted regressed LOG$_{10}$ median</th>
<th>BMUS formula gest age vs AFP</th>
<th>BMUS formula gest age vs LOG$_{10}$ (AFP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.382</td>
<td>0.385</td>
<td>0.385</td>
<td>0.354</td>
<td>0.389</td>
</tr>
<tr>
<td>2.5</td>
<td>0.469</td>
<td>0.474</td>
<td>0.473</td>
<td>0.422</td>
<td>0.463</td>
</tr>
<tr>
<td>5</td>
<td>0.531</td>
<td>0.537</td>
<td>0.537</td>
<td>0.486</td>
<td>0.532</td>
</tr>
<tr>
<td>10</td>
<td>0.618</td>
<td>0.616</td>
<td>0.617</td>
<td>0.554</td>
<td>0.608</td>
</tr>
<tr>
<td>15</td>
<td>0.676</td>
<td>0.680</td>
<td>0.683</td>
<td>0.617</td>
<td>0.673</td>
</tr>
<tr>
<td>20</td>
<td>0.735</td>
<td>0.726</td>
<td>0.726</td>
<td>0.662</td>
<td>0.724</td>
</tr>
<tr>
<td>25</td>
<td>0.781</td>
<td>0.785</td>
<td>0.770</td>
<td>0.702</td>
<td>0.764</td>
</tr>
<tr>
<td>30</td>
<td>0.824</td>
<td>0.821</td>
<td>0.821</td>
<td>0.741</td>
<td>0.812</td>
</tr>
<tr>
<td>35</td>
<td>0.862</td>
<td>0.853</td>
<td>0.854</td>
<td>0.778</td>
<td>0.848</td>
</tr>
<tr>
<td>40</td>
<td>0.906</td>
<td>0.911</td>
<td>0.911</td>
<td>0.817</td>
<td>0.896</td>
</tr>
<tr>
<td>45</td>
<td>0.966</td>
<td>0.950</td>
<td>0.947</td>
<td>0.869</td>
<td>0.950</td>
</tr>
<tr>
<td>50</td>
<td>1.000</td>
<td>1.001</td>
<td>1.002</td>
<td>0.901</td>
<td>0.987</td>
</tr>
<tr>
<td>55</td>
<td>1.050</td>
<td>1.048</td>
<td>1.054</td>
<td>0.952</td>
<td>1.042</td>
</tr>
<tr>
<td>60</td>
<td>1.100</td>
<td>1.105</td>
<td>1.105</td>
<td>1.006</td>
<td>1.098</td>
</tr>
<tr>
<td>65</td>
<td>1.156</td>
<td>1.169</td>
<td>1.168</td>
<td>1.054</td>
<td>1.152</td>
</tr>
<tr>
<td>70</td>
<td>1.235</td>
<td>1.232</td>
<td>1.231</td>
<td>1.107</td>
<td>1.210</td>
</tr>
<tr>
<td>75</td>
<td>1.313</td>
<td>1.303</td>
<td>1.310</td>
<td>1.186</td>
<td>1.295</td>
</tr>
<tr>
<td>80</td>
<td>1.400</td>
<td>1.390</td>
<td>1.395</td>
<td>1.263</td>
<td>1.375</td>
</tr>
<tr>
<td>85</td>
<td>1.500</td>
<td>1.516</td>
<td>1.515</td>
<td>1.365</td>
<td>1.495</td>
</tr>
<tr>
<td>90</td>
<td>1.688</td>
<td>1.674</td>
<td>1.680</td>
<td>1.517</td>
<td>1.656</td>
</tr>
<tr>
<td>95</td>
<td>1.941</td>
<td>1.958</td>
<td>1.957</td>
<td>1.765</td>
<td>1.931</td>
</tr>
<tr>
<td>97.5</td>
<td>2.250</td>
<td>2.242</td>
<td>2.250</td>
<td>2.023</td>
<td>2.212</td>
</tr>
<tr>
<td>99</td>
<td>2.625</td>
<td>2.643</td>
<td>2.647</td>
<td>2.381</td>
<td>2.600</td>
</tr>
</tbody>
</table>
Table 4.13
MoM distribution for 5082 consecutive HCG results. Distributions shown are MoMs derived using: weekly medians; weighted regressed weekly medians; weighted regressed LOG(weekly medians); regression of HCG concentration against BMUS formula derived gestation age; regression of LOG_{10}(HCG) concentration against BMUS formula derived gestation age.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Weekly median</th>
<th>Weighted regressed median</th>
<th>Weighted regressed LOG_{10} (median)</th>
<th>BMUS formula gest age vs HCG</th>
<th>BMUS formula gest age vs LOG_{10} (HCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.200</td>
<td>0.250</td>
<td>0.252</td>
<td>0.218</td>
<td>0.254</td>
</tr>
<tr>
<td>2.5</td>
<td>0.269</td>
<td>0.314</td>
<td>0.314</td>
<td>0.281</td>
<td>0.328</td>
</tr>
<tr>
<td>5</td>
<td>0.313</td>
<td>0.401</td>
<td>0.402</td>
<td>0.348</td>
<td>0.404</td>
</tr>
<tr>
<td>10</td>
<td>0.406</td>
<td>0.493</td>
<td>0.494</td>
<td>0.426</td>
<td>0.495</td>
</tr>
<tr>
<td>15</td>
<td>0.469</td>
<td>0.568</td>
<td>0.563</td>
<td>0.494</td>
<td>0.574</td>
</tr>
<tr>
<td>20</td>
<td>0.531</td>
<td>0.650</td>
<td>0.642</td>
<td>0.558</td>
<td>0.639</td>
</tr>
<tr>
<td>25</td>
<td>0.563</td>
<td>0.718</td>
<td>0.704</td>
<td>0.608</td>
<td>0.703</td>
</tr>
<tr>
<td>30</td>
<td>0.625</td>
<td>0.762</td>
<td>0.764</td>
<td>0.658</td>
<td>0.765</td>
</tr>
<tr>
<td>35</td>
<td>0.688</td>
<td>0.812</td>
<td>0.809</td>
<td>0.710</td>
<td>0.822</td>
</tr>
<tr>
<td>40</td>
<td>0.722</td>
<td>0.897</td>
<td>0.899</td>
<td>0.764</td>
<td>0.885</td>
</tr>
<tr>
<td>45</td>
<td>0.781</td>
<td>0.942</td>
<td>0.943</td>
<td>0.813</td>
<td>0.944</td>
</tr>
<tr>
<td>50</td>
<td>0.844</td>
<td>1.002</td>
<td>1.006</td>
<td>0.860</td>
<td>0.995</td>
</tr>
<tr>
<td>55</td>
<td>0.906</td>
<td>1.076</td>
<td>1.078</td>
<td>0.927</td>
<td>1.078</td>
</tr>
<tr>
<td>60</td>
<td>0.969</td>
<td>1.166</td>
<td>1.164</td>
<td>0.999</td>
<td>1.161</td>
</tr>
<tr>
<td>65</td>
<td>1.050</td>
<td>1.252</td>
<td>1.258</td>
<td>1.073</td>
<td>1.237</td>
</tr>
<tr>
<td>70</td>
<td>1.150</td>
<td>1.345</td>
<td>1.348</td>
<td>1.159</td>
<td>1.334</td>
</tr>
<tr>
<td>75</td>
<td>1.250</td>
<td>1.475</td>
<td>1.464</td>
<td>1.259</td>
<td>1.456</td>
</tr>
<tr>
<td>80</td>
<td>1.350</td>
<td>1.603</td>
<td>1.605</td>
<td>1.367</td>
<td>1.583</td>
</tr>
<tr>
<td>85</td>
<td>1.531</td>
<td>1.749</td>
<td>1.752</td>
<td>1.511</td>
<td>1.748</td>
</tr>
<tr>
<td>90</td>
<td>1.731</td>
<td>1.973</td>
<td>1.974</td>
<td>1.698</td>
<td>1.969</td>
</tr>
<tr>
<td>95</td>
<td>2.100</td>
<td>2.354</td>
<td>2.364</td>
<td>2.014</td>
<td>2.341</td>
</tr>
<tr>
<td>97.5</td>
<td>2.855</td>
<td>2.855</td>
<td>2.867</td>
<td>2.499</td>
<td>2.898</td>
</tr>
</tbody>
</table>
In the case of HCG MoMs, a similar situation is found. The weekly median derived MoM distribution and the BMUS-HCG median based MoM distribution are both significantly skewed relative to the other 3 distributions (approximate range 0.2 - 3.1 versus 0.25 - 3.45). Both of the skewed MoM distributions also have MoM = 1 at the 60th percentile. Thus, both weekly derived medians and medians derived from the regression of analyte concentration against BMUS formula derived gestation age are unsatisfactory for the derivation of MoMs because they do not result in a MoM distribution centred about 1.0. In conclusion therefore, the medians used for all further calculation of MoMs will be determined either by weighted regression of medians based on weekly medians or by regression of BMUS formula derived gestation age against LOG10(analyte concentration). The need to use LOG transformation before derivation of MoMs is not unexpected because the distribution of results at individual weeks of gestation (table 4.8) is non-Gaussian.

4.3.2 Effectiveness of Correction for Gestation Age

Conversion to MoMs is performed to remove the association between gestation age and analyte concentration. Therefore, if regression analysis of gestation age versus MoMs is performed, it is expected that there will be no correlation. For AFP MoM versus gestation age, the least squares regression parameters are: Slope, -0.00464; Intercept, 1.175; r, -0.00605; student’s t, -0.43 (Not significant). For HCG MoM versus gestation age, the least squares regression parameters are: Slope, 0.005628; Intercept, 1.06259; r, 0.00699; student’s t, 0.498 (Not significant). Both of these correlations confirm the lack of association between
gestation age and MoM and therefore that conversion to MoMs does effectively standardise for gestation age.

4.3.3 Normality of MoM distributions

Calculation of Down syndrome risks is based on Gaussian distribution functions and therefore to achieve meaningful results the population distribution of MoMs must conform to a Gaussian function. To assess the degree of conformity, the distributions of BMUS vs LOG transformed derived MoM in tables 4.12 and 4.13 and their LOGs were plotted on a probability scale (Barnett, 1979) (Figs 4.9 and 4.10). On a probability graph, normality is indicated by the points falling on a straight line. In both cases (AFP and HCG) raw MoMs form a curve but after conversion to LOG(MoM) the results lie on a straight line indicating that LOG transformation appears to render the population distribution Gaussian. This was further tested statistically by the Lilliefors test which resulted in test statistics of 0.021054 for the AFP MoM distribution and 0.011784 for the HCG MoM distribution (significance cut offs: \( \omega_{0.90} = 0.0113; \omega_{0.95} = 0.0124; \omega_{0.99} = 0.0145 \)). It must therefore be concluded that although Log\(_{10}\) transformation of AFP MoMs appears to be effective on examination of the probability plot, this is not true because the Lilliefors statistic exceeds both the 95% and 99% cut offs. The statistic for HCG lies between the 90% and the 95% cut offs which implies that normality cannot be excluded.
Figure 4.9.
Probability plot showing MoM distributions for AFP (squares) and LOG_{10}(AFP) (circles). Medians were derived by regression of the BMUS formula derived gestation dates against LOG_{10}(AFP).
Figure 4.10.
Probability plot showing MoM distributions for HCG (squares) and $\log_{10}(HCG)$ (circles). Medians were derived by regression of the BMUS formula derived gestation dates against $\log_{10}(HCG)$. 
4.3.4 Correlation of AFP MoMs and HCG MoMs

Since the calculation of Down syndrome risks currently relies on a multi-variate Gaussian distribution function, the correlation between different analytes is of great importance. The correlation coefficient for untransformed MoMs (AFP vs HCG) was 0.1544 (n = 5082, $P < 0.0001$) and for $\log_{10}$ transformed MoM was 0.1831 ($P << 0.0001$). These values are slightly different to the values for the unaffected population reported by Wald et al., 1988a who found a correlation between AFP and HCG MoMs of 0.05 in the unaffected population and 0.14 for the affected population using $\log$ transformed MoMs.
4.4 Conclusions

1) The distributions of AFP and HCG results are different at different weeks of gestation and are skewed and kurtosed with respect to a perfect Gaussian distribution (table 4.8). Therefore, LOG transformation must be performed if statistical manipulations relying on Gaussian assumptions are to be performed.

2) Derivation of medians is not simply a matter of choosing the median AFP or HCG concentration at each week. If this method is used, the distribution of MoMs will not necessarily conform to the ideal.

3) The ideal distribution for MoMs should be centred on MoM = 1. i.e. the point at which MoM = 1 should also be the 50th percentile of the distribution.

4) Week derived medians and medians derived by regressing raw AFP or HCG results against gestation date are not adequate. Either weekly medians should then be modified by weighted regression against gestation age or raw results should be LOG transformed and directly regressed against a measure of gestation age.

5) Conversion of raw serum results to MoMs effectively removes the relationship between the result and gestation age.

6) $\text{LOG}_{10}$ transformation of MoMs does not completely correct the population distribution of the MoMs to a perfect Gaussian distribution.
7) The correlation between AFP and HCG MoMs for the unaffected population is different to that previously reported.
chapter 5:

MATERNAL FACTORS AFFECTING MULTIPLES OF THE MEDIAN AND DOWN SYNDROME RISKS

5.1 MATERNAL Weight

5.1.1 The Rationale for Weight Correction
5.1.2 Preliminary Assessment of the Need for Weight Correction
5.1.3 The Effect of Weight correction in a Large Sample

5.2 MATERNAL HEIGHT

5.3 MATERNAL DIABETES

5.4 MATERNAL RACE

5.5 CONCLUSIONS
5.1 Maternal Weight

5.1.1 The Rationale for Weight Correction
(See also chapter 3: section 3.2.3)

Antenatal screening for Down’s syndrome using AFP, HCG and uE₃ was described in 1988 without reference to maternal weight correction (Wald et al, 1988). It had previously been observed that AFP is significantly and negatively correlated with maternal weight, and correction for weight of AFP values giving borderline positive results in neural tube defect (NTD) screening improved upon the false positive rate with negligible loss of detection (Wald et al, 1981). It is logical therefore, that if the rationale of weight correction of AFP at the high concentrations found in open NTD’s improves screening efficiency, then correction of AFP at the low levels expected for Down’s screening should also improve screening success. In addition, correction for maternal weight of HCG and uE₃ may also be applicable.

There has been debate about whether linear or exponential weight correction formulae are most appropriate (Haddow et al, 1981; Wald et al, 1981; Johnson and Lingley, 1984; Palomaki et al, 1985; Macri et al, 1986; Drugan et al, 1989; Johnson et al, 1990). The consensus appears to be that an exponential formula provides a better result because it works over a wider range of maternal weights. Therefore, the exponential method was used for all evaluations of maternal weight correction reported here.

5.1.2 Preliminary Assessment of the Need for Weight Correction
(Reynolds et al, 1991)

Over four months (Jan - Apr 1990 inclusive) all results from women being routinely screened for NTD’s by AFP assay and Down’s syndrome
by assay of AFP and HCG were examined. Gestational age was estimated by ultrasound scanning, or dates if scan unavailable. After exclusion of results of tests taken outside the gestational time window (15-19 weeks), from multiple pregnancies and from women without stated maternal age, there were 1840 results. Of these 1408 had a corresponding maternal weight (kg). A further 197 results were available with AFP, HCG and uE3. AFP was assayed using the in-house PEG separation RIA (RGH-AFP: chapter 3: section 3.1.1.1) and results were expressed as MoMs using weekly medians. The AFP medians for gestational dates 15-19 weeks inclusively were 26, 27, 33, 37, 37 kU/L respectively. HCG was assayed using the in-house magnetic separation IRMA (RGH-HCG: chapter 3: section 3.1.1.2) and results were expressed as MoM’s using weekly medians. The HCG medians for gestational dates 15-19 weeks were 30, 24, 18, 15, 14 U/mL respectively. Unconjugated oestriol was assayed using the Amerlex M 2nd trimester assay (IM2-uE3: chapter 3, section 3.1.3.3) and results were expressed as MoMs using weekly medians. The uE3 medians for gestational dates 15-19 weeks inclusively were 3.89, 4.47, 5.07, 6.45, 7.46 nmol/L respectively.

The mean weight of the 1408 women was 65.5 ± 8.75 (range 38 - 127 kg). Two women had been excluded because their weights exceeded 127 kg as this was above the range of the weigh-scales in the ante-natal clinic. For AFP and HCG a clear trend in median MoMs relative to weight was demonstrated. This was not found for uE3 (table 5.1). Regression parameters derived by least squares correlation between weight and $\log_{10}$ (AFP MoM), $\log_{10}$(HCG MoM) and $\log_{10}$uE3 MoM) are shown in table 5.2.
Table 5.1
Median serum AFP, HCG and uE₃ MoMs related to maternal weight.

<table>
<thead>
<tr>
<th>Maternal Weight (kg)</th>
<th>AFP MoM</th>
<th>HCG MoM</th>
<th>n</th>
<th>uE₃ MoM</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= 50</td>
<td>1.37</td>
<td>1.36</td>
<td>113</td>
<td>0.98</td>
<td>10</td>
</tr>
<tr>
<td>51 - 60</td>
<td>1.30</td>
<td>1.04</td>
<td>520</td>
<td>0.95</td>
<td>60</td>
</tr>
<tr>
<td>61 - 70</td>
<td>1.08</td>
<td>1.06</td>
<td>431</td>
<td>0.95</td>
<td>74</td>
</tr>
<tr>
<td>71 - 80</td>
<td>0.94</td>
<td>1.00</td>
<td>194</td>
<td>0.89</td>
<td>25</td>
</tr>
<tr>
<td>81 - 90</td>
<td>0.85</td>
<td>0.95</td>
<td>84</td>
<td>0.93</td>
<td>18</td>
</tr>
<tr>
<td>&gt;= 90</td>
<td>0.83</td>
<td>0.75</td>
<td>66</td>
<td>0.94</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 5.2
Regression parameters from least squares regression of weight against \( \log_{10}(\text{AFP MoM}) \), \( \log_{10}(\text{HCG MoM}) \) and \( \log_{10}(\text{uE}_3 \text{ MoM}) \).

<table>
<thead>
<tr>
<th></th>
<th>AFP MoM</th>
<th>HCG MoM</th>
<th>uE$_3$ MoM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.3890</td>
<td>0.2026</td>
<td>0.0144</td>
</tr>
<tr>
<td>B</td>
<td>-0.00532</td>
<td>-0.00288</td>
<td>-0.0010</td>
</tr>
<tr>
<td>95% confidence limits for B</td>
<td>±0.00079</td>
<td>±0.00095</td>
<td>±0.000186</td>
</tr>
<tr>
<td>r</td>
<td>-0.3315</td>
<td>-0.1563</td>
<td>-0.0758</td>
</tr>
<tr>
<td>n</td>
<td>1408</td>
<td>1408</td>
<td>197</td>
</tr>
<tr>
<td>Student’s t</td>
<td>13.177</td>
<td>5.933</td>
<td>1.06</td>
</tr>
<tr>
<td>P</td>
<td>&lt;&lt;0.001</td>
<td>&lt;&lt;0.001</td>
<td>&gt;0.1 (Not Sigt.)</td>
</tr>
</tbody>
</table>
Using the weight correction factors described in table 5.2, corrected MoMs were derived for all of the 1408 sets of data. This is achieved by calculating an expected MoM using maternal weight (e.g. a 60 kg patient would be expected to have an AFP MoM = $10^{(0.389 - (60 \times 0.00532))} = 1.17$). The corrected MoM is derived by dividing the observed MoM by the expected MoM. Down syndrome risks were derived for each of these sets of data using the parameters of Wald et al (1988a) and the likelihood ratio method (Reynolds and Penney, 1990). Four risks were derived from each set of data: a risk derived using uncorrected MoMs; a risk derived using weight corrected AFP and uncorrected HCG; a risk derived using weight corrected HCG and uncorrected AFP; and a risk using weight corrected AFP and HVG MoMs. At this time 5 of 7 cases of Down syndrome had been identified by the screening program in South Wales. The risks after correction were also evaluated in these 7 cases.

The effect of weight correction was determined by examining the screen positive rate for a risk cut off of 1:300 and by examining the cases for which correction altered the screening decision. Table 5.3 shows the effect of the different correction protocols. Correction of the Down syndrome case results did not change the overall detection rate but after correction of AFP and HCG, the risks assigned to the 2 missed cases had changed to be nearer to the risk cut off (1:380 became 1:320; 1:1140 became 1:930).

The effect of weight correction of AFP in NTD screening is well recognised (Wald et al, 1981, Johnson and Lingley, 1984). The proposed explanation for the relationship between weight and AFP is that heavier women have a larger volume of distribution when compared to that of
Table 5.3
The effect of weight correction on the screen positive rate in 1408 unaffected pregnancies. The figures in parentheses in columns 2 and 3 are the mean (± sd) weight of the women reclassified.

<table>
<thead>
<tr>
<th>Weight Correction</th>
<th>No. of cases in risk group (screen positive rate)</th>
<th>No. of cases reclassified as no longer at increased risk</th>
<th>No. of cases reclassified as now at increased risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>No correction</td>
<td>78 (5.5%)</td>
<td>4 (89 ± 7 kg)</td>
<td>17 (56 ± 6 kg)</td>
</tr>
<tr>
<td>AFP corrected</td>
<td>91 (6.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCG corrected</td>
<td>74 (5.3%)</td>
<td>9 (56 ± 6 kg)</td>
<td>5 (91 ± 13 kg)</td>
</tr>
<tr>
<td>AFP and HCG</td>
<td>82 (5.8%)</td>
<td>0</td>
<td>4 (55 ± 9 kg)</td>
</tr>
</tbody>
</table>
light women which results in a dilution effect on AFP. Although some workers have shown that it does not improve screening efficacy (Macri et al, 1986), the overall consensus is in favour of weight correction in NTD screening. Correction of weight in Down syndrome has also been advocated (Palomaki et al, 1985: Knight et al, 1988). Further indications of the importance of weight correction in Down syndrome include:

1) Wald and Cuckle, 1987 showed that there is no significant difference in weight between women delivering a Down syndrome baby and those delivering a normal baby.

2) When screening for Down syndrome using maternal age and AFP alone with no weight correction applied, the number of heavier women in the increased risk group is disproportionately high. Correction of AFP reduces this preponderance of heavier women (Knight et al, 1988).

Previous workers have failed to show any association between weight and HCG (Petrocik et al, 1989) but the number of patients studied was much smaller and despite showing a similar trend and r value (0.17) to the data described here, it was not statistically significant. A similar explanation to that used to explain the weight dependency of AFP is logical to explain the effect demonstrated for HCG. It is difficult to explain the lack of relationship between uE₃ and weight, although this may be due to the different processes involved in metabolism of steroid hormones, including protein binding, hepatic conjugation and excretion. The small sample size may also be implicated, although median uE₃ MoMs for different weight groups were remarkably consistent (table 5.1).
A further complicating factor is the weight correction method used. Palomaki et al, (1985) used a linear correction algorithm that has been demonstrated to be of use only over a limited weight range (45 - 90 kg) (Drugan et al, 1989). Other workers (Wald et al, 1981: Johnson and Lingley, 1984: Wald and Cuckle, 1987) recommend an exponential formula that results in a wider acceptable weight range (40 - 130 kg). The data presented above used the exponential method for the reason of the better weight range. If the woman is outside the weight range it has been recommended that correction should be applied to the nearest weight within the range (40 or 130 kg).

The effect of weight correction on Down syndrome risk screening is complex. Correction of AFP or HCG in a heavy woman results in a higher MoM value whilst in a light woman, the result is a lowering of the MoM value. The effect on the Down syndrome risk depends on the analyte corrected: An increase in the AFP MoM alone tends to result in a decrease in the estimated risk whilst an increase in the HCG MoM alone tends to result in an increase in the estimated risk. This effect can be clearly seen in the results in table 5.3 where 17 'light' women were reclassified as 'increased risk' after correction of AFP only and 5 'heavy' women were reclassified as at increased risk after correction of HCG only. A similar but opposite weight distribution was shown in women reclassified as 'not at increased risk'.

Correction of both AFP and HCG reduces the number of women reclassified because correction of one analyte partly compensates for the effect of correction of the other. Thus only 4 women were reclassified as 'increased risk' when correction of both analytes was performed and none
were reclassified as 'not at increased risk'. The mean weight of the reclassified women was low which is explained by the greater slope in the relationship between AFP and weight. It can be concluded that if weight correction is used, both or neither AFP and HCG must be corrected. It is not reasonable to correct only one analyte.

5.1.3 The Effect of Weight Correction in a Large Sample

Weight correction parameters for AFP and HCG were reassessed using data set 1 (comprising 5082 sets of results) as described in chapter 4, section 4.1.1 and for uE3 using a set of 495 cases. The median MoMs for different weight categories (as used in table 5.1) are shown in table 5.4. Least squares correlation of weight with LOG\text{10} transformed MoMs and the regression parameters are shown in table 5.5. Using these parameters, weight corrected MoMs were derived and the distribution of these compared to the the distribution of uncorrected MoMs is shown in table 5.6.

For AFP, it is evident that weight correction results in a tightening of the MoM distribution. For any pair of centiles straddling 50%, the inter-percentile range is less for weight corrected than for uncorrected MoMs. The effect for HCG is much smaller with a reduction in range of approximately 1% compared to an approximately 6% reduction in the range for AFP. This conforms the findings of Wald et al (1992a), who also showed a similar degree of narrowing of the AFP and HCG distributions after weight correction. The mean and standard deviation data for corrected and uncorrected LOG\text{10} transformed AFP and HCG MoMs for data set 1 and the ultrasound dated data for unaffected pregnancies from
Table 5.4. Median serum AFP, HCG and uE3 MoMs related to maternal weight in data set 1.

<table>
<thead>
<tr>
<th>Maternal Weight (kg)</th>
<th>AFP MoM</th>
<th>HCG MoM</th>
<th>n</th>
<th>uE3 MoM</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= 50</td>
<td>1.22</td>
<td>1.21</td>
<td>379</td>
<td>1.06</td>
<td>30</td>
</tr>
<tr>
<td>51 - 60</td>
<td>1.10</td>
<td>1.10</td>
<td>1763</td>
<td>1.06</td>
<td>171</td>
</tr>
<tr>
<td>61 - 70</td>
<td>0.96</td>
<td>0.98</td>
<td>1673</td>
<td>1.08</td>
<td>163</td>
</tr>
<tr>
<td>71 - 80</td>
<td>0.87</td>
<td>0.96</td>
<td>734</td>
<td>0.95</td>
<td>76</td>
</tr>
<tr>
<td>81 - 90</td>
<td>0.80</td>
<td>0.87</td>
<td>309</td>
<td>1.04</td>
<td>37</td>
</tr>
<tr>
<td>&gt;= 90</td>
<td>0.71</td>
<td>0.78</td>
<td>224</td>
<td>0.97</td>
<td>18</td>
</tr>
</tbody>
</table>
Table 5.5. Regression parameters for weight correction derived from data set 1 for AFP and HCG and from 495 cases with uE₃.

<table>
<thead>
<tr>
<th></th>
<th>AFP</th>
<th>HCG</th>
<th>uE₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (A)</td>
<td>0.3116</td>
<td>0.1935</td>
<td>0.0613</td>
</tr>
<tr>
<td>Slope (B)</td>
<td>-0.004804</td>
<td>-0.002983</td>
<td>-0.000854</td>
</tr>
<tr>
<td>95% confidence limits for B</td>
<td>± 0.00037</td>
<td>± 0.00052</td>
<td>± 0.0013</td>
</tr>
<tr>
<td>r</td>
<td>-0.3386</td>
<td>-0.1561</td>
<td>-0.0564</td>
</tr>
<tr>
<td>Student’s t</td>
<td>25.65</td>
<td>11.26</td>
<td>-1.25</td>
</tr>
<tr>
<td>P</td>
<td>&lt;&lt;0.001</td>
<td>&lt;&lt;0.001</td>
<td>0.21 (Not Significant)</td>
</tr>
</tbody>
</table>
Table 5.6.
Distribution of uncorrected AFP and HCG MoMs from table 4.12 (column 5) and table 4.13 (column 5) and weight corrected AFP and HCG MoMs (wAFP and wHCG).

<table>
<thead>
<tr>
<th>Percentile</th>
<th>AFP</th>
<th>wAFP</th>
<th>HCG</th>
<th>wHCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.389</td>
<td>0.426</td>
<td>0.254</td>
<td>0.250</td>
</tr>
<tr>
<td>2.5</td>
<td>0.463</td>
<td>0.494</td>
<td>0.328</td>
<td>0.329</td>
</tr>
<tr>
<td>5</td>
<td>0.532</td>
<td>0.553</td>
<td>0.404</td>
<td>0.407</td>
</tr>
<tr>
<td>10</td>
<td>0.608</td>
<td>0.629</td>
<td>0.495</td>
<td>0.498</td>
</tr>
<tr>
<td>15</td>
<td>0.673</td>
<td>0.688</td>
<td>0.574</td>
<td>0.577</td>
</tr>
<tr>
<td>20</td>
<td>0.724</td>
<td>0.738</td>
<td>0.639</td>
<td>0.643</td>
</tr>
<tr>
<td>25</td>
<td>0.764</td>
<td>0.781</td>
<td>0.703</td>
<td>0.709</td>
</tr>
<tr>
<td>30</td>
<td>0.812</td>
<td>0.822</td>
<td>0.765</td>
<td>0.773</td>
</tr>
<tr>
<td>35</td>
<td>0.848</td>
<td>0.861</td>
<td>0.822</td>
<td>0.823</td>
</tr>
<tr>
<td>40</td>
<td>0.896</td>
<td>0.923</td>
<td>0.885</td>
<td>0.892</td>
</tr>
<tr>
<td>45</td>
<td>0.950</td>
<td>0.946</td>
<td>0.944</td>
<td>0.950</td>
</tr>
<tr>
<td>50</td>
<td>0.987</td>
<td>0.988</td>
<td>0.995</td>
<td>1.013</td>
</tr>
<tr>
<td>55</td>
<td>1.042</td>
<td>1.034</td>
<td>1.078</td>
<td>1.076</td>
</tr>
<tr>
<td>60</td>
<td>1.098</td>
<td>1.081</td>
<td>1.161</td>
<td>1.155</td>
</tr>
<tr>
<td>65</td>
<td>1.152</td>
<td>1.136</td>
<td>1.237</td>
<td>1.235</td>
</tr>
<tr>
<td>70</td>
<td>1.210</td>
<td>1.196</td>
<td>1.334</td>
<td>1.334</td>
</tr>
<tr>
<td>75</td>
<td>1.295</td>
<td>1.276</td>
<td>1.456</td>
<td>1.447</td>
</tr>
<tr>
<td>80</td>
<td>1.375</td>
<td>1.356</td>
<td>1.583</td>
<td>1.559</td>
</tr>
<tr>
<td>85</td>
<td>1.495</td>
<td>1.469</td>
<td>1.748</td>
<td>1.725</td>
</tr>
<tr>
<td>90</td>
<td>1.656</td>
<td>1.609</td>
<td>1.969</td>
<td>1.936</td>
</tr>
<tr>
<td>95</td>
<td>1.931</td>
<td>1.869</td>
<td>2.341</td>
<td>2.337</td>
</tr>
<tr>
<td>97.5</td>
<td>2.212</td>
<td>2.120</td>
<td>2.898</td>
<td>2.845</td>
</tr>
<tr>
<td>99</td>
<td>2.600</td>
<td>2.501</td>
<td>3.461</td>
<td>3.470</td>
</tr>
</tbody>
</table>
Wald et al (1992a) are shown in table 5.7. It is noteworthy that the sd's for HCG are very similar in both studies but that the AFP sd's are much smaller in the Royal Gwent data. The effect of differences in population parameters on detection and false positive rates will be considered in chapter 8.

To assess the effect of weight correction on the false positive rate, Down syndrome risks were calculated from corrected and uncorrected AFP and HCG MoMs using the population parameters of Wald et al (1988a). Table 5.8 shows the numbers of 'screen positive' cases with risks from 1:100 to 1:500. It is noteworthy that there is a minimal change in the numbers of cases selected with either protocol and also that the results of this larger study confirm the preliminary study described in section 5.1.2. Thus for whatever risk cut off is chosen, weight correction can be applied without significantly affecting the numbers of women who should be advised to proceed to amniocentesis.

5.2 Maternal Height
(Reynolds et al 1992b)

Weight is not the only factor that may determine the distribution volume of AFP and HCG. If the distribution volume is determined by the total body water volume, a second factor that may be important is height because body water is related to the lean body mass. Thus in two women of equal weight, the shorter woman is likely to have a greater proportion of fat and therefore, a lesser body water volume. It may therefore be expected that even after correction for weight, shorter women will have a smaller distribution volume than taller women. The net effect of this would be that the median MoM in short women will be higher than in tall
Table 5.7.
Comparison of Gwent derived ‘unaffected’ population parameters for AFP and HCG with ‘unaffected’ population parameters reported by Wald et al (1992). Note: Wald et al assumed the mean MoMs for unaffected cases should be 1, hence LOG(MoM) = 0.

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data set 1</td>
<td>Wald</td>
</tr>
<tr>
<td><strong>AFP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>-0.00046</td>
<td>0.0</td>
</tr>
<tr>
<td>sd</td>
<td>0.1769</td>
<td>0.1910</td>
</tr>
<tr>
<td>n</td>
<td>5082</td>
<td>2028</td>
</tr>
<tr>
<td><strong>HCG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>-0.000032</td>
<td>0.0</td>
</tr>
<tr>
<td>sd</td>
<td>0.2386</td>
<td>0.2401</td>
</tr>
<tr>
<td>n</td>
<td>5082</td>
<td>2028</td>
</tr>
<tr>
<td>r (AFP v HCG)</td>
<td>0.1831</td>
<td>0.1749</td>
</tr>
</tbody>
</table>
Table 5.8.
Numbers of 'screen positive' results for uncorrected and weight corrected AFP and HCG MoMs from data set 1 (5082 cases).

<table>
<thead>
<tr>
<th>Risk cut off</th>
<th>Number of 'screen positive' cases</th>
<th>Uncorrected MoMs</th>
<th>Corrected MoMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:100</td>
<td>74 (1.45%)</td>
<td>73 (1.44%)</td>
<td></td>
</tr>
<tr>
<td>1:150</td>
<td>114 (2.24%)</td>
<td>114 (2.24%)</td>
<td></td>
</tr>
<tr>
<td>1:200</td>
<td>165 (3.25%)</td>
<td>160 (3.15%)</td>
<td></td>
</tr>
<tr>
<td>1:250</td>
<td>205 (4.03%)</td>
<td>203 (3.99%)</td>
<td></td>
</tr>
<tr>
<td>1:300</td>
<td>259 (5.09%)</td>
<td>256 (5.04%)</td>
<td></td>
</tr>
<tr>
<td>1:350</td>
<td>316 (6.22%)</td>
<td>316 (6.22%)</td>
<td></td>
</tr>
<tr>
<td>1:400</td>
<td>374 (7.36%)</td>
<td>369 (7.26%)</td>
<td></td>
</tr>
<tr>
<td>1:450</td>
<td>423 (8.32%)</td>
<td>420 (8.26%)</td>
<td></td>
</tr>
<tr>
<td>1:500</td>
<td>495 (9.74%)</td>
<td>488 (9.60%)</td>
<td></td>
</tr>
</tbody>
</table>
women. To investigate this, height was added to the special request form for the Royal Gwent Down screening program and a data set composed of 2857 cases with weight and height was collected (data set 2 as described in chapter 4, section 4.1). To determine whether height was significant and if so what was the best method for correction, a number of options were considered: correction for body mass index (weight / height²), correction for weight : height ratio (weight / height), correction for weight only and correction for weight and height using a multivariate correction. Initially correlation coefficients between $\log_{10}(AFP)$ and $\log_{10}(HCG)$ and the various height and weight combinations were derived (Table 5.9). Before weight correction, all weight and height combinations were significantly correlated with $\log_{10}$ transformed AFP and HCG MoMs. After weight correction, there were no significant correlations for HCG but for AFP there was still a correlation with height.

The effect of weight and height correction for AFP was further examined by deriving correction factors based on the weight and height combinations described above. The coefficients used are shown in table 5.10. MoMs corrected by each method were correlated against weight and height to determine which correction method was optimal. The correlation coefficients for each combination are shown in table 5.11. It is evident that correction for body mass index is ineffective because there are significant correlations between weight and height and body mass index corrected AFP. Similarly, a significant correlation exists between height and AFP corrected for weight : height ratio. The only protocol for correction that removes the influence of weight and height is correction for weight and height using the multivariate correction factor.
Table 5.9.
Correlations between weight and height and LOG_{10} transformed AFP and HCG MoMs (with and without weight correction). Total number of samples = 2857, therefore the two-tailed critical value for significance (P<= 0.05) is 0.03668.

<table>
<thead>
<tr>
<th></th>
<th>AFP</th>
<th>wAFP</th>
<th>HCG</th>
<th>wHCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>-0.387</td>
<td>0.000</td>
<td>-0.141</td>
<td>0.000</td>
</tr>
<tr>
<td>Height</td>
<td>-0.169</td>
<td>-0.058</td>
<td>-0.046</td>
<td>-0.004</td>
</tr>
<tr>
<td>Weight/Height</td>
<td>-0.367</td>
<td>0.012</td>
<td>-0.136</td>
<td>-0.001</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>-0.326</td>
<td>0.024</td>
<td>-0.125</td>
<td>-0.002</td>
</tr>
</tbody>
</table>
Table 5.10.
Coefficients used for weight and height correction. Correction factor derived as $10^{(A + B_{\text{parameter}})}$ for single parameter correction factors and as $10^{(A + B_{\text{weight}} + C_{\text{height}})}$ for multivariate weight and height correction.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate weight and height</td>
<td>0.53629</td>
<td>-0.0004758</td>
<td>-0.1515</td>
</tr>
<tr>
<td>weight / height</td>
<td>0.30055</td>
<td>-0.007959</td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.25959</td>
<td>-0.011256</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.11. Correlation coefficients between corrected MoMs and weight and height. Total number of samples = 2857, therefore the two-tailed critical value for significance (P<= 0.05) is 0.03668. wAFP = AFP MoM corrected for weight, whAFP = AFP MoM corrected for weight and height by multivariate method, w/hAFP = AFP MoM corrected for weight / height ratio. bAFP = AFP MoM corrected for body mass index.

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>whAFP</td>
<td>-0.0031</td>
<td>-0.0025</td>
</tr>
<tr>
<td>w/hAFP</td>
<td>0.0224</td>
<td>0.1239</td>
</tr>
<tr>
<td>bAFP</td>
<td>-0.0774</td>
<td>-0.1949</td>
</tr>
</tbody>
</table>
This result is not surprising. It would be remarkable if correction for body mass index compensated for weight and height because BMI is an index of obesity and it is possible for women of widely differing weights to have the same BMI. Since weight is the major contributor to the correlation for body size, removal of this factor will remove the effectiveness of correction. The failure of the weight : height ratio to correct effectively can be explained similarly.

The effect of multivariate weight and height correction on false positive rates was evaluated in the same way that weight correction was evaluated in section 5.1.3. Table 5.12 shows the number of 'screen positive' results with and without height correction of AFP. If these results are compared with those in table 5.8 it can be seen that the uncorrected numbers of 'screen positive' results differ significantly from the 'screen positive' rate in the larger data set. Since these rates are not comparable, it is only possible to compare the effect of weight and height correction with the effect of height correction. This reveals that there is little difference in the number of women selected by each approach. Unfortunately, no height data is available for Down syndrome associated pregnancies so the effect of height correction on false positive and detection rates cannot be examined in chapter 8 but the effect of weight correction will be considered.

5.3 Maternal Diabetes

It has been reported that pregnant women with diabetes mellitus tend to have lower serum AFP concentrations than women without diabetes (Wald et al, 1979; Milunsky et al, 1982; Reece et al, 1987; Green et
Table 5.12.
Numbers of 'screen positive' results for risks calculated using weight corrected AFP and HCG MoMs and for risks calculated using weight / height corrected AFP MoMs with weight corrected HCG MoMs from data set 2 (2857 cases).

<table>
<thead>
<tr>
<th>Risk cut off</th>
<th>Number of 'screen positive' cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uncorrected</td>
</tr>
<tr>
<td>1:100</td>
<td>25 (0.87%)</td>
</tr>
<tr>
<td>1:150</td>
<td>53 (1.86%)</td>
</tr>
<tr>
<td>1:200</td>
<td>87 (3.05%)</td>
</tr>
<tr>
<td>1:250</td>
<td>128 (4.48%)</td>
</tr>
<tr>
<td>1:300</td>
<td>156 (5.46%)</td>
</tr>
<tr>
<td>1:350</td>
<td>198 (6.93%)</td>
</tr>
<tr>
<td>1:400</td>
<td>229 (8.02%)</td>
</tr>
<tr>
<td>1:450</td>
<td>256 (8.96%)</td>
</tr>
<tr>
<td>1:500</td>
<td>294 (10.3%)</td>
</tr>
</tbody>
</table>
al, 1988). This has recently also been shown for uE₃ and HCG and confirmed for AFP (Wald et al, 1992b). The medians reported in diabetes are: AFP, 0.77; HCG, 0.95; and uE₃, 0.92. However, the reports of decreased levels of AFP have also recently been questioned by Zimmermann et al, (1992) who found no significant decrease in AFP levels in 53 women with gestational diabetes when compared to a group of 560 normal women.

In order to test the hypothesis that diabetes causes alterations in the concentration of AFP and HCG, the database of the Royal Gwent Hospital screening program was interrogated and all of the results with maternal diabetes flagged were identified. This revealed 31 women with known diabetes mellitus. Before examining the AFP and HCG results of these patients, their ages and weights were compared by the Smirnov test (chapter 3, section 3.4.4) with the age and weight distributions of the Gwent pregnant population (tables 4.1 and 4.2 respectively) to ensure that as far as possible, factors other than diabetes were not responsible for any differences that may be found.

The Smirnov maximum vertical difference for the age distributions was 0.152 (critical value, \( \omega_{0.95} = 0.245 \)) and for the weight distribution was 0.443 with the same critical value. This proved that the ages of the diabetic women were not atypical of the population but that the weights were significantly different. The mean population weight was 64.9 ± 12.4 kg and the mean weight of the diabetics was 66.9 ± 9.1 kg. This difference is very small but nevertheless statistically significant. Therefore, the effect of diabetes was evaluated on both uncorrected and weight corrected multiples of the median for AFP and HCG.
The Smirnov maximum vertical distance for uncorrected AFP was 0.279 and for weight corrected AFP, 0.247 (critical value, $\omega_{0.95} = 0.245$). Both of these distances are statistically significant, although in the case of weight corrected AFP this is only just significant. The mean AFP MoM for the diabetics was 0.79 (0.81, weight corrected) compared to 1.00 for the normal population. For HCG the distances were 0.125 (uncorrected) and 0.096 (weight corrected), neither of which are statistically different.

Although the number of cases of diabetes available for study was small, it can tentatively be concluded that the previous reports of decreased AFP in diabetics are correct; the recent report of slightly decreased HCG (Wald et al., 1992b) was based on 126 normal and 92 diabetic pregnancies and found that the decrease was not statistically significant ($P=0.48$). This small study therefore agrees with previously published results.

5.4 Maternal Race

Maternal race has been postulated to affect AFP levels in pregnancy. Proposed differences have included decreased AFP in Indians (Shapiro et al., 1975) and increased AFP in American blacks (Crandall et al., 1983; Johnson, 1985; Baumgarten, 1986). Racial differences are however controversial as other workers have shown no difference due to race in American blacks (Macri et al., 1976; Milunsky et al., 1979). To test the conclusions of these earlier publications, the data base of the Gwent screening program was interrogated to find results for non-caucasian women. This revealed 36 sets of results which were analysed for age, weight AFP and HCG as described for diabetic results in section 5.3.
The Smirnov maximum vertical distance for the age distribution was 0.137 (critical value, $\omega_{0.95} = 0.227$) and for the weight distribution was 0.054 (critical value, $\omega_{0.95} = 0.241$) indicating that the ages and weights of the non-caucasian women were not significantly different from the total population. Since no difference was found for the weight distributions, only uncorrected AFP and HCG MoMs were examined. For AFP, the Smirnov maximum vertical distance was 0.215 and for HCG, 0.087 (critical value, $\omega_{0.95} = 0.227$). Neither of these results are statistically significant. However, this study is far too small to determine conclusively whether there are significant differences between the caucasian and non-caucasian populations.
5.5 Conclusions

1) There is a highly significant relationship between weight and analyte concentration for AFP and HCG but not for uE3. This relationship remains after conversion of analyte concentration to MoMs.

2) After correction for maternal weight using a weight correction factor derived using the exponential algorithm, the weight relationship is removed.

3) There is also a significant correlation between height and AFP MoM that is still present after correction for weight.

4) Correction for height and weight can be achieved using a multi-variate regression derived correction factor. Correction for body mass index or weight / height ratio is not effective.

5) Weight correction or weight / height correction does not significantly affect the 'screen positive' rate. The effect on detection and false positive rates of weight correction will be examined in chapter 8.

6) The AFP levels in diabetic women are decreased relative to the levels in non-diabetic women. The mean MoM in a group of 31 diabetics was 0.79 which agrees well with previously published results. There is no significant difference for HCG.

7) No significant differences were demonstrated between caucasian and non-caucasian populations.
chapter 6:

**THE IMPORTANCE OF ACCURATE GESTATION DATING**

**IN DOWN SYNDROME SCREENING**

### 6.1 THE PROBLEM OF GESTATION DATING

### 6.2 DEMONSTRATION OF ULTRASOUND IMPRECISION

### 6.3 DEMONSTRATION OF THE EFFECT OF ULTRASOUND IMPRECISION ON SCREENING

6.3.1 Demonstration of Edge Effects due to Ultra-Sonic Dating

6.3.2 Demonstration of 'Quantum' Effects due to Ultra-sonic dating

### 6.4 CONCLUSIONS
6.1 THE PROBLEM OF GESTATION DATING  
(Reynolds, 1992: Reynolds et al, 1992)

Gestation dating is crucial to the calculation of meaningful Down syndrome risk factors because the concentrations of the biochemical parameters change as gestation progresses. Thus, the result for each analyte must be standardised to a value which is independent of gestational age. Currently this is achieved by calculation of a multiple of the median (MoM) where the laboratory result is divided by the median result of a large group of women at a similar stage of gestation. This must be previously determined from the results of a large number of women with proven gestation dates. Therefore, if an accurate MoM is to be determined an accurate gestational age is needed.

The way in which gestation dates are expressed may also have an influence on the Down syndrome risk screening calculation. It is common practice in many hospitals to express gestation dates in whole weeks (integer weeks), either by rounding LMP dates to the nearest week or by truncation. Thus a woman who has a gestation of 16 weeks and 4 days may be regarded as 16 weeks in one unit and 17 weeks in another. Also, where dates are assessed by ultrasound they are often presented as whole weeks and sometimes as ‘16+’ or ‘17+’. This is particularly difficult to interpret when deciding which median to use.

There are several methods used to determine gestational age: palpation of uterine size; calculation from the date of last menstrual period (LMP dating); estimation of fetal age by measurement of crown-rump length (CRL) or biparietal diameter (BPD). None of these methods are perfect and all have a spectrum of error (Table 6.1). LMP dating is
particularly poor for dating because of variation in menstrual cycle length, failure to correctly remember the date of the period, missing light periods etc., resulting in the potential for a ± 4 week error.

Since LMP dating is known to be a poor indicator in many cases, let us consider ultrasound dating. This may be achieved by measuring a foetal dimension and comparing it against a size chart (table 6.2) to determine gestation week or by entering the dimension in a regression equation to determine gestation week ± day as described in chapter 3, section 3.5. It must be remembered however that ultrasonic measurement is the same as a biochemical assay in that it has inherent imprecision and that there is biological variation between fetuses of similar age. It has been shown for axial measurements by ultrasound (e.g. BPD) that there is an intra-observer error of ± 1mm and an inter-observer variation of ± 1.3 mm for a 51 mm spherical test object in a water bath (Pretorius et al, 1984). This corresponds to CV’s of 2.0% and 2.5% respectively. However, this was achieved under standard test conditions with specially calibrated equipment so the precision achieved in routine practice is likely to be worse than this.

6.2 Demonstration Of Ultrasound Imprecision

Figure 6.1 shows a Gaussian distribution curve. If the horizontal axis represents the ‘true BPD’ (ie. the exact fetal head size assuming no error at all), the line below represents a ‘Week size band’. For the sake of example this could be the 16 week size band which ranges from 34 - 38 mm (Table 6.2). In this example the vertical line corresponds to a ‘true’ BPD size of 35.4 mm. Any measurement error is assumed to be normally
Table 6.1
Accuracy of clinical and ultrasound estimation of gestational age (Reed, 1990).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Trimester</th>
<th>Variation</th>
<th>Estimated CV at 16 weeks gestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pelvic Exam</td>
<td>1</td>
<td>± 7 days</td>
<td>N/A</td>
</tr>
<tr>
<td>LMP</td>
<td>1,2,3</td>
<td>± 3 weeks</td>
<td>9.4%</td>
</tr>
<tr>
<td>CRL</td>
<td>1</td>
<td>± 5 days</td>
<td>N/A</td>
</tr>
<tr>
<td>BPD</td>
<td>1,2</td>
<td>± 10 days</td>
<td>4.5%</td>
</tr>
<tr>
<td>Femur</td>
<td>1,2</td>
<td>± 10 days</td>
<td>4.5%</td>
</tr>
</tbody>
</table>
Table 6.2. Biparietal diameter size bands corresponding to gestation week as used at the Royal Gwent until mid-1991 when they were replaced by the BMUS formula method of deriving gestation dates.

<table>
<thead>
<tr>
<th>gestation age (weeks)</th>
<th>BPD size band (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>&gt;26.0 - 30.0</td>
</tr>
<tr>
<td>15</td>
<td>&gt;30.0 - 34.0</td>
</tr>
<tr>
<td>16</td>
<td>&gt;34.0 - 38.0</td>
</tr>
<tr>
<td>17</td>
<td>&gt;38.0 - 41.5</td>
</tr>
<tr>
<td>18</td>
<td>&gt;41.5 - 45.0</td>
</tr>
<tr>
<td>19</td>
<td>&gt;45.0 - 49.0</td>
</tr>
</tbody>
</table>
Figure 6.1.
Diagram to show how the probability of an incorrect gestation week is derived.
distributed about the ‘true BPD’ value and therefore the Gaussian distribution curve spreads out on either side of the vertical line dependent on the CV. If gestation week is determined on the basis of size bands, then there is a chance that the measured BPD will be in the band higher or lower than the true size band. The magnitude of this chance ($P_{\text{incorrect-low}}$ or $P_{\text{incorrect-high}}$) is determined by calculating the number of standard deviations the size band edge is away from the ‘true BPD’. The corresponding probability for each standard deviation can be read from a Gaussian distribution table and by adding $P_{\text{incorrect-low}}$ and $P_{\text{incorrect-high}}$ probabilities, the probability of an incorrect week assessment is derived.

Figure 6.2 demonstrates the effect of imprecision on the confidence which can be placed on an integer week assessment by ultrasound. The graph clearly shows that it is very difficult to accurately determine integer gestation week by ultrasound especially if the foetal head size is close to a size-band interface. This is because any inaccuracy in BPD determination is amplified by the conversion of the relatively continuously distributed BPD result into a discontinuous ‘discrete’ variable (week). If a less discrete value (eg. week + day or the BPD value itself) is used, the peaks and troughs in figure 6.2 would be less marked and the effects of gestation date estimation on the calculated risk factors would be minimised.

6.3 Demonstration Of The Effect Of Ultrasound Imprecision On Screening

The effect of gestation dating imprecision may be demonstrated in two ways, by edge effects and by computer simulation of 'quantum' effects. Firstly, edge effects will be considered:
Figure 6.2.
Probability of incorrect gestation week assignment for BPD assessed pregnancy according to 'true' BPD of fetus. The size bands used to determine gestation week are shown in table 6.2 and the bands for 16 and 17 weeks are shown.
6.3.1 Edge Effects in Ultrasound Dating Errors

The screening data of 4605 women routinely screened for Down syndrome associated pregnancy between Jan and Nov 1990 were examined. After exclusion of 1060 data sets from women with multiple pregnancies, women who had not been scanned and incomplete data sets, and exclusion of 523 results from women who had not been scanned on the same day that the blood sample was collected for screening, there were 3022 complete sets of data. A complete data set included age at expected date of delivery, BPD (measured outer table to outer table), AFP and HCG.

Table 6.2 shows the size bands used to determine gestation week. Many laboratories receive ultrasound gestation age as a decimal value (eg. 16.1, 15.5). These values are derived from a similar size band table or by calculation using regression parameters for BPD versus dates (see chapter 3, section 3.5). Although they appear to be more accurate they still suffer the same uncertainty as other ultrasound derived measurements as described in section 6.1.

Log transformation of AFP and HCG values results in normalisation of the data and therefore, least squares regression of \( \log_{10}(AFP) \) and \( \log_{10}(HCG) \) against raw BPD was performed to derive medians for AFP or HCG relative to BPD (equivalent to BMUS derived medians, as chapter 4: section 4.2.4). Previous workers (Wald et al, 1988a) have derived regressed medians by deriving the median for each week individually, followed by weighted least squares regression. Medians for weeks 13 - 27 were used and, for example, at 16 weeks there were 122
patients. Thus, to calculate the regression, the value for the 16 week median was entered into the least squares regression 122 times. For many of the weeks used, medians were based on less than 10 patient results. The regressed medians used here are based on the whole data set and are statistically valid because Log transformation results in a Gaussian distribution for the data. Furthermore, only data for weeks 14 - 20 was used to generate regression parameters and regressed medians were only considered valid for weeks 15-19 (minimum data set was for 15 weeks, n=228). The regression parameters are shown in table 6.3.

With these regression parameters it is possible to calculate regressed medians for BPD derived gestation week based and BPD based gestation dating methods (eg for a 16 week pregnancy, the regressed median for AFP = 10\(^{(0.7495+(16x0.047339))}\) = 32.1 kU/l). Regressed medians for gestation weeks 15 - 18 and for extremes of BPD size bands are shown in table 6.4. In this table, the BPD size band extremes are the points on a BPD size band at which a minimal change in measured BPD (0.1 mm) is sufficient to change the estimated week of gestation (eg 34.0 mm corresponds to 15 weeks and 34.1 mm to 16 weeks).

To examine the effect on screening of varying the method of calculating the MoMs, a series of values were tested using the medians shown in table 6.4. These were derived from gestation week or extremes of BPD size band using the regression parameters in table 6.3. AFP and HCG values were chosen such that using week based medians, the MoMs would be 0.5, 1.5 and 2.5 assuming gestation week = 16. Using these values, MoMs using BPD derived medians assuming BPD was at each extreme of the size band were calculated. These values are shown in
Table 6.3
Exponential regression coefficients for gestation week or raw BPD (mm) against $\text{LOG}_{10}(\text{AFP})$ (kU/l) or $\text{LOG}_{10}(\text{HCG})$ (U/ml).

Formula for regressed mean $= 10^{(A + Bx)}$
where $x =$ week or BPD size, $A$ and $B =$ regression coefficients

<table>
<thead>
<tr>
<th>Gestation week</th>
<th>BPD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AFP</td>
</tr>
<tr>
<td>A</td>
<td>0.7495</td>
</tr>
<tr>
<td>B</td>
<td>0.047339</td>
</tr>
<tr>
<td>r</td>
<td>0.2252</td>
</tr>
</tbody>
</table>
Table 6.4.
Calculated medians for each week of gestation and for extremes of BPD size bands. Medians calculated using formula shown in legend to table 6.3.

<table>
<thead>
<tr>
<th>Week</th>
<th>AFP (kU/l)</th>
<th>HCG (U/ml)</th>
<th>Medians based on biparietal diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Medians based on gestation week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Medians based on gestation week</td>
</tr>
<tr>
<td>15</td>
<td>28.8</td>
<td>26.1</td>
<td>30.1  25.8  29.8</td>
</tr>
<tr>
<td>16</td>
<td>32.1</td>
<td>22.9</td>
<td>34.0  29.5  25.4</td>
</tr>
<tr>
<td>17</td>
<td>35.8</td>
<td>20.0</td>
<td>34.1  29.6  25.3</td>
</tr>
<tr>
<td>18</td>
<td>40.0</td>
<td>17.5</td>
<td>38.0  33.7  21.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>38.1  33.8  21.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>41.5  37.8  18.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>41.6  38.0  18.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>45.0  42.5  16.2</td>
</tr>
</tbody>
</table>
The MoMs calculated were applied to the multivariate screening algorithm (Reynolds and Penney, 1990) using parameters previously described (Wald et al, 1988a). The Down syndrome risk factors calculated using these MoMs and the percentage differences between the gestation week based MoM Down syndrome risk factor and the BPD based MoM Down syndrome risk factors are shown in table 6.6.

Finally, the effect of a 0.1mm change in BPD resulting in a change from one size band to another was examined. Risks were calculated using week based means to calculate MoMs for values corresponding to 0.5, 1.5 and 2.5 MoMs for BPD equal to 38 and 38.1 mm. These MoMs are shown in table 6.7 and the week based MoM Down syndrome risk factor and percentage difference between risk assuming gestation week = 16 and gestation week = 17 are shown in table 6.8.

### 6.3.1.1 The Significance of Edge Effects

The wide variation in MoMs that can occur due to gestation dating errors is a function of the variation in concentration of the markers in maternal serum throughout pregnancy and some method of assessing gestational age must be employed to allow standardisation of the measured concentrations. The original description of Down syndrome screening using the multivariate method used gestational dating by LMP (Wald et al, 1988a). This method is problematical because of poor recall of date of LMP and variations in cycle length. Similarly, estimation of gestational age by ultrasound may be rendered inaccurate due to variations in fetal growth rate and poor recall of LMP which makes collection of accurate data to evaluate the relationship between BPD and
Table 6.5.
BPD based MoMs calculated for AFP and HCG values corresponding to week based MoMs of 0.5, 1.5 and 2.5 for a 16 week gestation pregnancy.

<table>
<thead>
<tr>
<th>Week based MoM</th>
<th>BPD based MoM</th>
<th>BPD based MoM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AFP</td>
<td>HCG</td>
</tr>
<tr>
<td></td>
<td>BPD=34.1</td>
<td>BPD=38.0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.542</td>
<td>0.476</td>
</tr>
<tr>
<td>1.5</td>
<td>1.627</td>
<td>1.429</td>
</tr>
<tr>
<td>2.5</td>
<td>2.711</td>
<td>2.381</td>
</tr>
</tbody>
</table>

%age difference from week based MoM

+ 8.4%   - 6.5%   - 9.4%   + 6%
Table 6.6.
Risks calculated for a woman of 16 weeks gestation who will be 20 years old at expected date of delivery. The MoM quoted in the table legend refers to the MoM for week based medians. MoMs used for BPD$_{\text{low}}$ and BPD$_{\text{high}}$ are those in table 6.5. BPD$_{\text{low}}$ = 34.1 mm. BPD$_{\text{high}}$ = 38.0 mm

Risk factors shown are such that risk = 1:n.
Factors are shown to 6 significant figures to emphasise differences. Risks generated in screening would not be reported to this accuracy.

<table>
<thead>
<tr>
<th>HCG MoM</th>
<th>0.5 MoM</th>
<th>1.5MoM</th>
<th>2.5MoM</th>
</tr>
</thead>
<tbody>
<tr>
<td>risk</td>
<td>%diff</td>
<td>risk</td>
<td>%diff</td>
</tr>
<tr>
<td></td>
<td></td>
<td>risk</td>
<td>%diff</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP MoM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 MoM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPD$_{\text{low}}$</td>
<td>6386</td>
<td>858</td>
<td>266</td>
</tr>
<tr>
<td>week</td>
<td>4629</td>
<td>600</td>
<td>183</td>
</tr>
<tr>
<td>BPD$_{\text{high}}$</td>
<td>3814</td>
<td>484</td>
<td>146</td>
</tr>
<tr>
<td>1.5 MoM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPD$_{\text{low}}$</td>
<td>64664</td>
<td>5994</td>
<td>1565</td>
</tr>
<tr>
<td>week</td>
<td>45465</td>
<td>4062</td>
<td>1043</td>
</tr>
<tr>
<td>BPD$_{\text{high}}$</td>
<td>36873</td>
<td>3228</td>
<td>821</td>
</tr>
<tr>
<td>2.5MoM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPD$_{\text{low}}$</td>
<td>185914</td>
<td>14502</td>
<td>3496</td>
</tr>
<tr>
<td>week</td>
<td>129017</td>
<td>9699</td>
<td>2299</td>
</tr>
<tr>
<td>BPD$_{\text{high}}$</td>
<td>103745</td>
<td>7643</td>
<td>1794</td>
</tr>
</tbody>
</table>
Table 6.7.
Week based MoMs calculated for AFP and HCG values corresponding to BPD based MoMs of 0.5, 1.5 and 2.5 for a foetus with BPD = 38.0 - 38.1mm

<table>
<thead>
<tr>
<th>BPD based MoM</th>
<th>AFP week=16</th>
<th>AFP week=17</th>
<th>HCG week=16</th>
<th>HCG week=17</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.525</td>
<td>0.471</td>
<td>0.472</td>
<td>0.54</td>
</tr>
<tr>
<td>1.5</td>
<td>1.575</td>
<td>1.412</td>
<td>1.415</td>
<td>1.62</td>
</tr>
<tr>
<td>2.5</td>
<td>2.625</td>
<td>2.353</td>
<td>2.358</td>
<td>2.70</td>
</tr>
</tbody>
</table>

%age difference from BPD based MoM: +5% - 5.8% - 5.6% + 8%
Table 6.8.
Risks calculated for a woman with BPD=38.0-38.1 who will be 20 years old at expected date of delivery. The MoM quoted in the table legend refers to the MoM for BPD based medians. MoMs used for week 16 and 17 are those in table 6.7.

Risk factors are such that risk = 1:n. Factors are shown to 6 significant figures to emphasis differences. Risks generated in screening would not be reported to this accuracy.

<table>
<thead>
<tr>
<th>HCG MoM</th>
<th>0.5 MoM risk</th>
<th>%diff</th>
<th>1.5 MoM risk</th>
<th>%diff</th>
<th>2.5 MoM risk</th>
<th>%diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP MoM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 MoM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week=16</td>
<td>5607</td>
<td>-35.4%</td>
<td>742</td>
<td>-38.5%</td>
<td>229</td>
<td>-40.2%</td>
</tr>
<tr>
<td>week=17</td>
<td>3623</td>
<td>-35.4%</td>
<td>456</td>
<td>-38.5%</td>
<td>137</td>
<td>-40.2%</td>
</tr>
<tr>
<td>1.5 MoM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week=16</td>
<td>5607</td>
<td>-38.0%</td>
<td>5114</td>
<td>-41.0%</td>
<td>1326</td>
<td>-42.3%</td>
</tr>
<tr>
<td>week=17</td>
<td>34727</td>
<td>-38.0%</td>
<td>3017</td>
<td>-41.0%</td>
<td>765</td>
<td>-42.3%</td>
</tr>
<tr>
<td>2.5 MoM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week=16</td>
<td>160195</td>
<td>-39.2%</td>
<td>12308</td>
<td>-42.1%</td>
<td>2946</td>
<td>-43.4%</td>
</tr>
<tr>
<td>week=17</td>
<td>97476</td>
<td>-39.2%</td>
<td>7125</td>
<td>-42.1%</td>
<td>1667</td>
<td>-43.4%</td>
</tr>
</tbody>
</table>
gestational age difficult. This inability to accurately date a pregnancy by LMP or BPD may partially explain previous reports that BPD dating does not improve screening efficiency in AFP screening for Down syndrome (Cuckle and Wald, 1987) whilst combined LMP and BPD dating does offer some improvement in neural tube defect screening (Roberts et al, 1979). Further complications include: numerous different protocols for relating BPD to gestation week; differences in the fetal growth rate between races; and differences in methods of measuring BPD (inner to outer table, outer to outer table) (Campbell and Newman, 1971; Sabbagh et al, 1976; Wiener et al, 1977; Sabbagha and Hagle, 1978; Parker et al, 1982; Munoz et al, 1986; Cuckle and Wald, 1987).

The investigation described in section 6.3.1 avoided many of the above pitfalls because a single BPD : age conversion protocol was used, all measurements of BPD were by the same method (outer to outer table) and the population of Gwent is mostly racially homogeneous. The results in table 6.6 clearly demonstrate that if a risk is calculated using the two variables (AFP and HCG) and MoMs derived from gestation week estimated by ultra-sound, the risk generated may be significantly at variance with the risk factor calculated using BPD derived medians. Thus if the fetal head is at the lowest end of the size band it is possible that the risk calculated is too high by up to 45% which may result in an unnecessary referral for amniocentesis. Similarly if the fetal head is at the top end of the size band the risk calculated may be up to 22% too low, resulting in a failure to refer for amniocentesis.

Table 6.8 provides further evidence that basing MoM calculation on gestation week derived from ultra-sound causes major inaccuracies. A
change in BPD of only 0.1mm resulting in a change of size band is sufficient to change the calculated risk by 40%. These results only cover the effect of using two screening variables. The effect of using three variables in the screen can be estimated by conservatively assuming a ± 5% variation between the uE3 week based and BPD based medians and applying the resulting values to the risk calculation algorithm as before. This results in an estimate that using three variables and gestation week based medians, the error may be 55-150% for fetal head sizes at the lowest end of the size band and 25-60% at the top of the size band.

Furthermore, these estimates of errors in risk do not take account of analytical errors. The error for ultrasound measurement of BPD has been reported as approximately ± 1.5 mm (Lunt and Chard, 1974) which at a BPD of 34mm corresponds to 4.5%. Pretorius et al, (1984) also showed significant imprecision in ultrasound measurements. Newer ultrasound machines may be more accurate but as the above studies did not include observer error and measured spherical test objects under ideal conditions, this is unlikely. The unavoidable errors from analytical variation in the biochemical parameters can only serve to worsen the situation. Deliberate loss of precision by converting a relatively precise measure of gestation (BPD) to an imprecise measure (gestation week) can and should be avoided. It is therefore logical that if gestation dates are to be checked by ultrasound then medians based either on BPD or gestation day should be used instead of medians based on gestation week derived from BPD.

As a caution however, it should be remembered that genetically abnormal fetuses may have abnormally small or large heads. Obviously if the defect is gross it will be recognised and appropriate action taken but
minor abnormalities may not be detected. This could affect gestation staging either by BPD alone or BPD derived gestation week.

6.3.2 Demonstration of 'Quantum' Effects due to Ultra-Sonic Dating

Having demonstrated the large errors that may result in a single specimen due to edge effects, it is necessary to examine the result of ultrasound imprecision on a population. Since it is not possible to perform enough repeat ultrasound measurements on a single woman to measure the effect of imprecision on a large sample, the only option is to use computer simulation. A simulation of the effect of gestation dating errors for a 'double screen' (AFP + HCG) was performed by setting target values for BPD (34 and 36 mm) which were modified using the random normal variate method (for CV = 3, 4, 5 and 6%) as described in chapter 3, section 3.3.1. Only AFP + HCG screening was simulated because simulation of AFP only screening is pointless as this is poorly effective in Down screening and because data directly relating uE₃ to BPD was not available preventing simulation of the 'triple screen'. Integer gestation week was derived by comparison with the size bands in table 6.2. The integer gestation week or the raw ‘simulated BPD’ was then used to determine the median used to calculate the MoM. Medians derived directly from BPD were calculated using the regression parameters in table 6.3. No analytical variation was applied to the biochemical variables which were set as: AFP, 0.72 MoM, 22.3 kU/L; HCG, 2.04 MoM, 49 U/mL. The simulation was run for 30,000 cases and likelihood ratios were calculated from both the integer week derived and BPD derived MoMs.

Tables 6.9 and 6.10 show the distribution of likelihood ratios
Table 6.9.
Distribution of likelihood ratios for different ultrasound measurement imprecision (CV’s 3 - 6%) using integer week derived MoMs and BPD target = 34mm. Value given is frequency (as %) of likelihood ratio in 30,000 simulated events.

<table>
<thead>
<tr>
<th>Likelihood Ratio</th>
<th>Assumed CV for Ultrasound Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>0.122</td>
<td>0.0</td>
</tr>
<tr>
<td>0.358</td>
<td>45.7</td>
</tr>
<tr>
<td>0.672</td>
<td>54.3</td>
</tr>
<tr>
<td>1.2721</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table 6.10.
Distribution of likelihood ratios for different ultrasound measurement imprecision (CV's 3 - 6%) using integer week derived MoMs and BPD target = 36mm. Value given is frequency (as %) of likelihood ratio in 30,000 simulated events.

<table>
<thead>
<tr>
<th>Likelihood Ratio</th>
<th>Assumed CV for Ultrasound Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>0.072</td>
<td>0.0</td>
</tr>
<tr>
<td>0.122</td>
<td>2.6</td>
</tr>
<tr>
<td>0.358</td>
<td>93.9</td>
</tr>
<tr>
<td>0.672</td>
<td>3.5</td>
</tr>
<tr>
<td>1.2721</td>
<td>0.0</td>
</tr>
</tbody>
</table>
calculated using integer week MoMs for BPD = 34mm and 36mm respectively. It can be seen that using integer gestation week derived medians, only a limited number of likelihood ratios are possible. These individual possibilities each represent a single 'week' decision, i.e. likelihood ratio = 0.122 is generated when the simulated BPD indicates a gestation age = 17 weeks, likelihood ratio = 0.672 is generated when the simulated BPD = 15 weeks. The frequencies of each likelihood ratio are related to the BPD target. Thus in table 6.10, most likelihood ratios correspond to gestation week = week 16 but in table 6.9 there is a significant number of likelihood ratios corresponding to week = 15. It is also important to note that the values of likelihood ratios are the same regardless of the BPD target.

Tables 6.11 and 6.12 show the distribution of likelihood ratios when MoMs are derived directly from BPD without conversion to dates. It can be seen that there is a spread of results forming a distribution instead of a small number of discrete result quanta. The spread of the distribution of likelihood ratios increases as the imprecision of ultrasound increases but the ratios derived using BPD derived medians form a much tighter distribution than when weekly derived medians are used. It is also evident that the modal likelihood ratio differs depending on the BPD target (the modal likelihood ratio for BPD target = 34 mm is 0.40 - 0.45 whilst for a BPD target = 36 mm, the modal likelihood ratio is 0.30 - 0.35).

To more effectively compare the effect of dating method, the results from tables 6.9 and 6.11 and from tables 6.10 and 6.12 are presented graphically as figures 6.3 and 6.4 respectively. The graphs show that if integer gestation weeks are used, the multi-modal distribution of
Table 6.11
Distribution of likelihood ratios for different ultrasound measurement imprecision (CV’s 3 - 6%) using BPD derived MoMs for BPD target = 34mm. Value given is frequency (as %) of likelihood ratio in 30,000 simulated events.

<table>
<thead>
<tr>
<th>Likelihood Ratio</th>
<th>Assumed CV for Ultrasound Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>0.15 - 0.20</td>
<td></td>
</tr>
<tr>
<td>0.20 - 0.25</td>
<td>0.12</td>
</tr>
<tr>
<td>0.25 - 0.30</td>
<td>0.36</td>
</tr>
<tr>
<td>0.30 - 0.35</td>
<td>4.68</td>
</tr>
<tr>
<td>0.35 - 0.40</td>
<td>15.79</td>
</tr>
<tr>
<td>0.40 - 0.45</td>
<td>30.00</td>
</tr>
<tr>
<td>0.45 - 0.50</td>
<td>22.17</td>
</tr>
<tr>
<td>0.50 - 0.55</td>
<td>15.88</td>
</tr>
<tr>
<td>0.55 - 0.60</td>
<td>7.60</td>
</tr>
<tr>
<td>0.60 - 0.65</td>
<td>2.52</td>
</tr>
<tr>
<td>0.65 - 0.70</td>
<td>0.86</td>
</tr>
<tr>
<td>0.70 - 0.75</td>
<td>0.14</td>
</tr>
<tr>
<td>0.75 - 0.80</td>
<td>0.52</td>
</tr>
<tr>
<td>0.80 - 0.85</td>
<td>0.20</td>
</tr>
<tr>
<td>0.85 - 0.90</td>
<td>0.29</td>
</tr>
<tr>
<td>0.90 - 0.95</td>
<td>0.19</td>
</tr>
<tr>
<td>0.95 - 1.00</td>
<td>0.09</td>
</tr>
<tr>
<td>1.00 - 1.05</td>
<td>0.16</td>
</tr>
<tr>
<td>1.05 - 1.10</td>
<td>0.26</td>
</tr>
<tr>
<td>1.10 - 1.15</td>
<td>0.06</td>
</tr>
<tr>
<td>1.15 - 1.20</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Table 6.12
Distribution of likelihood ratios for different ultrasound measurement imprecision (CV's 3 - 6%) using BPD derived MoMs for BPD target = 36mm. Value given is frequency (as %) of likelihood ratio in 30,000 simulated events.

<table>
<thead>
<tr>
<th>Likelihood Ratio</th>
<th>Assumed CV for Ultrasound Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>0.10 - 0.15</td>
<td></td>
</tr>
<tr>
<td>0.15 - 0.20</td>
<td>3.91</td>
</tr>
<tr>
<td>0.20 - 0.25</td>
<td>23.01</td>
</tr>
<tr>
<td>0.25 - 0.30</td>
<td>35.57</td>
</tr>
<tr>
<td>0.30 - 0.35</td>
<td>24.37</td>
</tr>
<tr>
<td>0.35 - 0.40</td>
<td>9.65</td>
</tr>
<tr>
<td>0.40 - 0.45</td>
<td>2.95</td>
</tr>
<tr>
<td>0.45 - 0.50</td>
<td>0.54</td>
</tr>
<tr>
<td>0.50 - 0.55</td>
<td>0.61</td>
</tr>
<tr>
<td>0.55 - 0.60</td>
<td>0.38</td>
</tr>
<tr>
<td>0.60 - 0.65</td>
<td>0.49</td>
</tr>
<tr>
<td>0.65 - 0.70</td>
<td>0.17</td>
</tr>
<tr>
<td>0.70 - 0.75</td>
<td>0.03</td>
</tr>
<tr>
<td>0.75 - 0.80</td>
<td>0.26</td>
</tr>
<tr>
<td>0.80 - 0.85</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Figure 6.3.
Comparison of likelihood ratios calculated using MoMs derived from weekly medians and BPD-derived medians for a simulated data set with target BPD = 34 mm. Circles = likelihood ratios derived from week derived MoMs, Triangles = likelihood ratios derived from BPD derived MoMs.
Figure 6.4.
Comparison of likelihood ratios calculated using MoMs derived from weekly medians and BPD-derived medians for a simulated data set with target BPD = 36 mm. Circles = likelihood ratios derived from week derived MoMs, Triangles = likelihood ratios derived from BPD derived MoMs.
likelihood ratios generated results in a significant proportion of events being away from the major peak. BPD-derived gestation dating shows spread of results but the distribution is continuous and is narrower than the multi-modal distribution. This clearly shows that expression of gestation age in integer weeks is unsatisfactory because ‘quantum’ effects create a multi-modal risk profile. It can be concluded that gestational age must be quoted in smaller units than weeks. Either BPD measured on the day of sample collection could be provided to the laboratory or if this is not possible (for instance units which perform booking scans with estimation of CRL) then date must be provided in weeks and days. This could mean that laboratories need their own medians for each day of gestation. Obviously this would be difficult to achieve because it would entail a vast sample size to allow each median to be determined individually. Fortunately the same effect can be achieved by least squares regression of the gestational age measurement against \( \text{LOG}_{10} \) transformed AFP or HCG data (Reynolds et al, 1992) (see also chapter 4, section 4.2).
6.4 Conclusions

1) Accurate gestation dating is vital if accurate Down syndrome risks are to be estimated. Therefore confirmation of LMP dates is preferable if not mandatory.

2) Ultrasound measurements have imprecision just as do biochemical assays. Ultrasound dating may therefore result in the wrong median being chosen for calculation of the MoM.

3) Expression of dates in integer weeks can lead to massive changes in estimated risks due to a small change in BPD measurement.

4) Gestation dates must be expressed in days to avoid the multi-modal 'quantum' effect.
chapter 7:

LABORATORY INFLUENCES IN DOWN SYNDROME SCREENING

7.1 THE IMPORTANCE OF ASSAY OPTIMISATION

7.1.1 Comparison of a poorly optimised assay (IM2-uE3 Canick / Wald modification) with a fully optimised assay (IM4-uE3)
7.1.1.1 Comparison of Assay Standardisation
7.1.1.2 Comparison of Assay Results
7.1.1.3 The Effect of Conversion to MoMs
7.1.1.4 The Effect on Likelihood Ratios
7.1.1.5 The Effect on Assay Imprecision
7.1.1.6 Conclusions
7.1.2 Comparison of two optimised AFP assays

7.2 THE EFFECT OF ASSAY IMPRECISION

7.3 COMBINATION OF ASSAY AND DATING IMPRECISION

7.4 CONCLUSIONS
7.1 THE IMPORTANCE OF ASSAY OPTIMISATION

7.1.1 Comparison of a poorly optimised assay (IM2-uE3 Canick / Wald modification) with a fully optimised assay (IM4-uE3) (Reynolds and John, 1992)

In all of the seminal papers on Down syndrome screening in which uE3 was measured (Canick et al, 1988; Wald et al, 1988a; Wald et al, 1988b), and in work which questioned its usefulness (Macri et al, 1990a), the assay utilised was a ‘third trimester’ assay kit (either the IM2 series Amerlex oestriol RIA kit or the IM3 series Amerlex-M specific unconjugated oestriol kit) which had been empirically modified without re-optimisation by increasing the sample volume and diluting the low standards to increase sensitivity at low concentrations of oestriol (chapter 3, section 3.1.3). These kits have now been superseded by a new assay kit specifically optimised for the oestriol concentrations found in the second trimester (the IM4 series Amerlex-M second trimester oestriol kit) (chapter 3, section 3.1.3).

Screening for Down syndrome in pregnancy arose from neural tube defect screening programs when it was noticed that maternal serum alpha-fetoprotein (AFP) concentrations were decreased in Down syndrome affected pregnancies. The multiple of the median (MoM) was used in neural tube screening (Wald and Cuckle, 1977; Wald and Cuckle, 1980) and became accepted for calculation of Down risk factors when these were described in 1984 (Merkatz et al, 1984). Parvin et al (1991) demonstrated that significant inter-assay differences may result from conversion of results to MoMs. To evaluate the importance of assay optimisation and determine the significance of the effect described by Parvin, the IM4-uE₃ kit (chapter 3, section 3.1.3.3) and the IM2-uE₃
Canick/Wald method (chapter 3, section 3.1.3.5) were compared.

Serum sent for routine Down syndrome screening at the UHW was used. All samples were collected from patients who had their gestation dates assessed by ultrasound examination. There were 61 samples from patients at 15 weeks gestation, 57 at 16 weeks and 59 at 17 weeks. All samples were assayed by both the IM4-uE₃ and the IM2-uE₃-Canick/Wald methods. Quality control samples and standards from both kits were also analysed by both methods. All QC samples gave results within the manufacturer’s recommended ranges when assayed with the kit they were supplied for. Coefficients of variation for each method were assessed from duplicate results by the method of Ekins (1983) using a running calculation from more than 220 samples. Comparisons between methods were all performed using Deming’s regression with variance ratio = 1. Likelihood ratios and Down syndrome risk estimates were calculated using the multivariate Gaussian method (chapter 3, section 3.2.4: Reynolds and Penney, 1990).

7.1.1.1 Comparison of Assay Standardisation

The counts for the IM4-uE₃ standard curve (expressed as B/B₀) using the ‘optimised’ IM4-uE₃ assay plotted against uE₃ concentration are shown in Fig 7.1. Also plotted on this graph are the counts for the standards from the IM2-uE₃ Canick/Wald method assayed by the IM4-uE₃ assay. The counts for the IM2-uE₃ Canick/Wald standard curve (expressed as B/B₀) assayed using the ‘modified’ IM2-uE₃ Canick/Wald assay plotted against uE₃ concentration are shown in Fig 7.2. Also plotted on this graph are the counts for the standards from the IM4-uE₃
Figure 7.1.
Dose response curves for IM4-uE₃ and IM2-uE₃ standards assayed using the IM4-uE₃ assay. Triangles represent results for standards from the IM4-uE₃ kit and circles represent the results for standards from the IM2-uE₃ Canick / Wald modified assay.
Figure 7.2.
Dose response curves for IM4-uE3 and IM2-uE3 standards assayed using the IM2-uE3 Canick / Wald assay. Triangles represent results for standards from the IM4-uE3 kit and circles represent the results for standards from the IM2-uE3 Canick / Wald modified assay.
method assayed by the IM2-uE₃ Canick / Wald assay. In figure 7.2 it can be seen that when assayed with the IM2-uE₃ Canick / Wald method, the 5 nmol/L oestriol standard from the IM4-uE₃ kit gives approximately the same percentage binding of $^{125}$I-oestriol as the 2 nmol/L standard from the IM2-uE₃ Canick / Wald assay. The results for the zero standards cannot be shown on a log scale but the IM4 kit 2.2 nmol/L standard had the same percentage binding as the IM2 kit zero standard. The effect of this is that the zero, 1.1 and 2.2 nmol/L oestriol standards from the IM4 kit assayed using the IM2-uE₃ Canick / Wald method, give oestriol concentrations which are less than zero.

The ‘zero’ oestriol standard in the IM2-uE₃ kit was a pooled human serum and the extra standards prepared in the Canick / Wald method were prepared using this zero oestriol standard. The zero oestriol standard in the IM4-uE₃ kit is prepared by a different method that ensures that all oestriol is removed (C. Davies, Kodak Diagnostics Ltd., personal communication). It can therefore be inferred that the sub-zero oestriol concentrations in the IM4-uE₃ standards indicate that there was some oestriol present in the IM2-uE₃ zero standard. Further evidence that the IM2-uE₃ assay standards contain more estriol than their stated concentration is shown in figures 7.1 and 7.2: In both in cases, the standards from the IM4-uE₃ assay displace less $^{125}$I-oestriol from their respective antibodies than the standards from the IM2-uE₃ assay.

7.1.1.2 Comparison of Assay Results

The expected result of the IM2-uE₃ Canick / Wald method’s standards containing an increased oestriol concentration in comparison with their assigned concentration is that patient samples should give
results that are lower than their 'true' values when assayed by the
IM2-uE₃ Canick / Wald modified procedure using modified IM2-uE₃
standards. This was observed when the IM4-uE₃ kit standards were used
in the IM2-uE₃ kit to standardise the assay of patient samples, i.e the
patient results were higher. However, when the results of assaying a
sample with both kits are compared, a different picture is seen: patient
samples assayed using the IM4-uE₃ kit, have consistently lower results
than when they are assayed with the IM2-uE₃ Canick / Wald assay.
Regression parameters comparing uE₃ concentrations assayed using the
IM2-uE₃ kit and the IM4-uE₃ kit were calculated using Deming’s
regression as: Intercept = - 5.8, Slope = 1.89 (95% confidence limits for
slope = 1.74 - 2.07), r = 0.864, n = 177. Figure 7.3 is a scattergram
showing the uE₃ concentrations with the Deming’s regression line and
95% confidence limits.

The reason why this should be is not obvious as the IM4-uE₃ kit
oestriol standards produced the expected change in patient results when
used in the IM2-uE₃ assay, as described above. This apparent
inconsistency is probably because the Canick / Wald modification
resulted in alterations in assay kinetics, matrix effects and non-specific
binding. The utilisation of a different antibody in the IM4-uE₃ kit may
also have contributed to the differences in results with the new assay.

7.1.1.3 The Effect of Conversion to MoMs

Medians for each week for both assays were determined from the
patient results. These were: IM2-uE₃ kit: 15 weeks, 4.52 nmol/L; 16
weeks, 4.67 nmol/L; 17 weeks, 5.34 nmol/L: IM4-uE₃ kit: 15 weeks, 2.07
Figure 7.3.
Comparison of uE₃ concentrations in patient samples assayed using the IM2-uE₃ Canick / Wald modified kit and the IM4-uE₃ kit. The solid line is the Deming’s regression line, the dashed lines are the 95% confidence limits.
nmol/L; 16 weeks, 3.44 nmol/L; 17 weeks, 4.4 nmol/L. Using these medians, MoMs for all results were calculated and IM2-uE₃ kit MoMs were compared with IM4-uE₃ MoMs by Deming’s regression. The regression parameters were: Intercept = -1.3, Slope 2.32 (95% confidence limit for slope = 2.14 - 2.53), r = 0.8717, n = 177. Figure 7.4 is a scattergram of the MoMs derived from the unconjugated estriol results with Deming’s regression line and 95% confidence limits.

The MoM frequency distributions for each assay were determined and are shown in table 7.1. Both distributions have the same mean value (1.0) but a different spread. Comparison of the cumulative frequency distributions by the Smirnov test (chapter 3, section 3.4.4) showed that the two distributions were different (max diff = 0.2203, critical value for 95% confidence = 0.1445).

### 7.1.1.4 The Effect on Likelihood Ratios and Down Syndrome Risks

To evaluate the effect of the increased spread of MoMs using the optimised assay, the likelihood ratios for each sample were calculated using the AFP and HCG MoMs already determined and used for screening, and the uE₃ MoMs derived from IM2-uE₃ or IM4-uE₃ results. Regression parameters calculated for LOG₁₀ transformed data comparing likelihood ratios calculated using each set of uE₃ MoMs were: A = -0.252; B = 1.625; r = 0.8522 (fig 7.5).

The Down syndrome risk factors using both sets of estriol MoMs were calculated using the age-related risk derived from the age at expected date of delivery for the woman the sample was collected from.
Figure 7.4. Scattergram with Deming's regression line and 95% confidence limits comparing MoMs for patient samples assayed using the IM2-uE₃ assay and the IM4-uE₃ assay. The solid line is the Deming's regression line, the dashed lines are the 95% confidence limits.
Table 7.1.
Distribution of MoMs for IM2 and IM4 kits.

<table>
<thead>
<tr>
<th>%ile</th>
<th>IM2 kit</th>
<th>IM4 kit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>0.71</td>
<td>0.35</td>
</tr>
<tr>
<td>5.0</td>
<td>0.78</td>
<td>0.47</td>
</tr>
<tr>
<td>10.0</td>
<td>0.83</td>
<td>0.58</td>
</tr>
<tr>
<td>25.0</td>
<td>0.89</td>
<td>0.75</td>
</tr>
<tr>
<td>50.0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>75.0</td>
<td>1.15</td>
<td>1.36</td>
</tr>
<tr>
<td>90.0</td>
<td>1.27</td>
<td>1.70</td>
</tr>
<tr>
<td>95.0</td>
<td>1.35</td>
<td>1.86</td>
</tr>
<tr>
<td>97.5</td>
<td>1.45</td>
<td>2.11</td>
</tr>
</tbody>
</table>
Figure 7.5
Scattergram comparing likelihood ratios calculated from IM2 and IM4 kit derived uE₃ MoMs. The solid line is the Deming's regression line, the short dashed lines are the 95% confidence limits, the long dashed line is the line of equality.
The screen positive samples were identified using a risk cut off of 1:300. Using the MoMs derived from the concentrations assayed by the IM2-uE₃ kit the screen positive rate was 4.5% (8 samples) whilst using the IM4-uE₃ kit the screen positive rate was 6.8% (12 samples). The significance of this was tested by the $\chi^2$ test and found to be not significant ($\chi^2 = 0.85$).

7.1.1.5 The Effect on Assay Imprecision

The precision profiles of the two assays are shown in Fig 7.6. The CV of the IM4-uE₃ optimised assay is less than 10% down to estriol concentrations of 1 nmol/L whereas with the IM2-uE₃ modified assay the coefficient of variation is greater than 10% at estriol concentrations below 4 nmol/L.

7.1.1.6 Conclusions

In order to measure the lower serum uE₃ concentrations in the second trimester, Canick et al (1988) modified the oestriol assay normally used for measuring estriol in third trimester samples. This modification involved doubling the sample volume to 40 $\mu$L and including two extra standards prepared by diluting the lowest estriol standard with the zero standard supplied with the kit. However, no data were presented substantiating the modification.

Prior to conversion of absolute values to MoMs it was believed that this would reduce the apparent differences between the two sets of uE₃ results. However, the differences were enhanced by conversion to MoMs. This creates difficulties in Down syndrome risk factor calculation.
Figure 7.6.
Intra-assay precision profiles for uE₃ assays.
Circles represent IM2-uE₃ Canick / Wald modified assay.
Triangles represent IM4-uE₃ assay.
The effect of different assays on likelihood ratios (fig 7.5) clearly shows that using MoMs derived from the IM4-uE₃ kit results, a wider spread of likelihood ratios is calculated than if the IM2-uE₃ assay results are used to derive MoMs.

Clearly, the inadequate optimisation of the IM2-uE₃ modification and the introduction of new assay reagents and standards in the IM4-uE₃ kit has introduced a methodological bias between the two assays. Since the original oestriol kit on which the Down risk estimates were calculated is no longer available and patient results assayed with the IM4-uE₃ 'optimised' procedure are lower than with the IM2-uE₃ 'modified' assay, then new population parameters are required to ensure that risk estimates derived with the new oestriol assay are valid.

With the data available and the discontinuation of the IM2-uE₃ kit, it is impossible to estimate the effect of the change in the assay on the Down syndrome detection rate, but evaluation of the screen positive rate (equivalent to false positive rate) showed that a greater proportion of women would be identified for amniocentesis if the IM4-uE₃ kit results were used. Obviously the number of samples which gave screen positive results in this study was very small and thus the lack of statistical significance of the increase from 4.5% to 6.8% screen positives is not surprising but if the increase were shown to be a real effect by a larger study, it would have significant financial implications due to the increase in amniocenteses and karyotyping which would be required. However it would be possible to reduce the increase in screen positives by changing the parameters used in the risk calculation algorithm, or by changing the risk cut off.
Changing the parameters used in the multivariate calculation of Down syndrome risk to avoid the effect of a change in the MoM distribution would be problematical because it would be necessary to calculate population parameters for every assay used by every laboratory for both normal and Down syndrome pregnancies. This would cause few problems for the ‘normal’ group but since the incidence of Down syndrome is relatively low, it would take many years before individuals laboratories could develop their own values for the Down syndrome population. Also, it would mean that changes in assays would be impossible as the entire standardisation process would have to be restarted. This would be impossible as even a change in the specificity of antibodies could have a major effect on results. Similarly, altering the local risk cut-off value would be fraught with difficulty.

A solution to the problem of inter-assay variability in MoMs is not easy to find but may include developing a ‘world-standard’ for the frequency distribution of MoMs. In order for this to work, a suitable assay would be used to define a ‘normal’ (not necessarily Gaussian) MoM distribution function with which all assays should be compared. If the MoM frequency distribution for an assay differs from this ideal distribution, the MoMs would need to be adjusted until they conform to the ‘world standard’. In the example here, a simple linear transformation of \( \log_{10} \) transformed MoMs could be used to convert MoMs from the new assay to an equivalent value for the old assay. A more attractive possibility would be to avoid the use of MoMs altogether but this is not easy to achieve.

In conclusion, the comparison of the two assays for unconjugated
estriol revealed a significant difference in the oestriol concentrations measured due to a methodological bias. This inevitably means that the median values for estriol at each week of gestation for each assay are different. It is also significant that conversion of the results to MoMs does not remove the inter-assay differences. This confirms the work of Parvin et al (1991). The result is a difference in the likelihood ratios calculated using MoMs from different assays which for the 177 samples analysed gave screen positive rates of 4.5% (IM2-uE3 assay) and 6.8% (IM4-uE3 assay). Since the published parameters for uE3 in Down syndrome risk calculation (Wald et al, 1988a) are based on the old assay which is no longer available, either new parameters must be derived or a new method of standardising for gestational age which does not use MoMs is vital.

7.1.2 Comparison of two optimised AFP assays

Two hundred and sixty samples from women with gestations between 15 and 19 weeks were assayed with both the UHW-AFP assay and the Amersham second trimester AFP assay. The AFP results were then converted to MoMs. Demings regression was used to determine the relationship between AFP results and AFP MoMs. For AFP results, regression parameters were: Intercept = 0.029; Slope = 1.088 (95% confidence limits for slope = 0.986 - 1.202), \( r = 0.779 \). For AFP MoMs, regression parameters were: Intercept = -0.0075; Slope = 1.022 (95% Confidence limits for slope = 0.925 - 1.128), \( r = 0.776 \). These results (i.e. the fact that the slope and intercept are extremely near to 1 and 0 respectively) demonstrate that the AFP results and MoMs are much closer in agreement than the uE3 results and MoMs were. This is further shown in table 7.2, where it is evident that the MoM distribution for AFP
Table 7.2.
Distribution of MoMs for UHW-AFP and Amersham AFP assays.

<table>
<thead>
<tr>
<th></th>
<th>UHW-AFP</th>
<th>Amersham</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 %ile</td>
<td>0.52</td>
<td>0.35</td>
</tr>
<tr>
<td>5.0 %ile</td>
<td>0.60</td>
<td>0.57</td>
</tr>
<tr>
<td>10.0 %ile</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>25.0 %ile</td>
<td>0.77</td>
<td>0.82</td>
</tr>
<tr>
<td>50.0 %ile</td>
<td>0.98</td>
<td>1.02</td>
</tr>
<tr>
<td>75.0 %ile</td>
<td>1.27</td>
<td>1.26</td>
</tr>
<tr>
<td>90.0 %ile</td>
<td>1.53</td>
<td>1.56</td>
</tr>
<tr>
<td>95.0 %ile</td>
<td>1.90</td>
<td>1.86</td>
</tr>
<tr>
<td>97.5 %ile</td>
<td>2.06</td>
<td>2.09</td>
</tr>
</tbody>
</table>
measured with either assay is almost identical compared to the marked differences in distribution for uE₃ (Table 7.1).

It can therefore be concluded that a correctly optimised assay is likely to compare well with another correctly optimised assay but that empirically modified assays that have not been proven such as the IM2-uE₃ Canick / Wald modification do not perform acceptably when compared to an assay that has been specifically designed for its analyte.

7.2 THE EFFECT OF ASSAY IMPRECISION
(Reynolds, 1992)

The likelihood ratio method for calculation of Down syndrome risk is an effective method of combining a number of laboratory results into a single numerical value. In theory, by increasing the spread of possible results and accentuating differences between normal and Down syndrome results, the addition of extra analytes into the risk calculation improves the detection rate whilst maintaining the false positive rate constant (Spencer, 1991). Thus, addition of uE₃ to AFP and HCG results should provide a more effective screening test. This has been borne out by a number of trials: Mancini et al (1991), Nørgaard-Pedersen et al (1990) and Amerlite Diagnostics (1991) concluded that the ‘triple screen’ should be used although the benefits of adding uE₃ were small. However, Macri et al (1990a, 1990b, 1990e), Spencer et al (1992) and Crossley et al (1993) believe that a ‘double test’ utilising AFP + HCG (or free β-HCG) is sufficient and that addition of uE₃ is counter-productive.

In all laboratory tests, there is variation due to inter- and intra-
individual factors and analytical factors (imprecision). These analytical factors are critical because they influence the ability of the test to discriminate between ‘normal’ and ‘abnormal’ groups. If imprecision becomes too great, the separation between groups becomes blurred and the test loses its effectiveness. Imprecision may be evaluated by examining the coefficient of variation (CV) of the assay. This is derived by making repeated measurements on a single sample and calculating the mean and standard deviation of the results. The CV is the standard deviation divided by the mean, multiplied by 100 to express the results as a percentage.

When more than one test result is combined into a single parameter, the imprecision of each result contributes to the total imprecision of the parameter. It is difficult to accurately calculate the CV for Down syndrome risk factors because of the complexity of the mathematical algorithm required to calculate the risk estimate. However, it is possible to estimate CVs for laboratory imprecision in Down syndrome risk estimates by repeated assay or by computer simulation (chapter 3; section 3.3).

To assess the imprecision in the risk estimate, 4 samples from women of 16 weeks gestation were assayed for AFP and HCG in replicate (n = 22) using the UHW-AFP and UHW-HCG assays. The results of each analysis were used to derive likelihood ratios. The Down syndrome risk factor (likelihood ratio x Age risk) was not calculated since the same age risk would have been used for each set of results. For the ‘1-parameter’ (AFP only) screen, 22 likelihood ratios were derived and for the ‘2-parameter’ (AFP + HCG) screen, 484 ratios were derived. This was
achieved by combining each AFP value with all of the HCG values in the calculation. This approach is valid because any of the combinations could have occurred in a ‘real’ sample. The CV of each set of likelihood ratios was then calculated. The mean of the 4 imprecision estimates was taken as a representative estimate of the CV of Down syndrome risk estimates. For AFP only risk estimates, the CV was 12.3% and for AFP + HCG risk estimates, the CV was 24.7%.

Imprecision was also estimated by computer simulation using the methods described in chapter 3, section 3.3. Analytical CVs for the AFP, HCG and uE3 were obtained from Amerlite kit insert data-sheets. Target levels for each analyte were set to correspond to the medians expected in Down syndrome as described by Wald et al (1988a). The levels chosen were: AFP: 0.72 MoM, 22.3 kU/l, CV = 5.8%; HCG: 2.04 MoM, 49 U/ml, CV = 5.1%; and uE3: 0.73 MoM, 3.26 nmol/l, CV = 8.7%. Analytical variation was simulated on these values and when each set of simulated results was complete they were converted into MoMs and the likelihood ratio was calculated using the population parameters of Wald et al (1988a). The simulation was run for 10,000 cycles before CVs were calculated. The CVs for the simulated analysis were: For AFP only risk estimates, 8.6%; for AFP + HCG risk estimates, 15.7%; and for AFP + HCG + uE3 risk estimates, 23.8%.

The CV of ‘3-parameter’ risk factors assessed by simulation is similar to CVs previously reported (21.4 - 47.8%) (Holding, 1991). Also it is significantly higher than the CVs for 1 or 2-parameter risk factors. This is unsurprising given the increased number of parameters required in the 3-parameter likelihood ratio calculation algorithm. It may be inferred
that increasing imprecision was behind the failure of attempts to further improve the Down screen by addition of a fourth variable (pregnancy specific β1-glycoprotein) (Wald et al, 1989).

Analytical imprecision has long been considered important in clinical chemistry. In 1963, it was proposed that the maximum coefficient of variation allowable for any assay should be 10% (Tonks, 1963), although this was later amended to 20% for certain analytes (Tonks 1968). Furthermore, surveys of physicians have determined that the ‘medically useful’ CV is also very close to 20% and that CVs greater than this are unacceptable (Elion-Gerritzen, 1980; Skendzel et al, 1985). Another method of expressing the acceptability limits for precision is that analytical imprecision should be equal to or less than one half of the average within subject variability (i.e. the intrinsic fluctuation about the homeostatic setting point of an individual) (Young et al, 1971; Harris E, 1979; W.A.S.P. 1979; Fraser, 1990).

A CV of 20% means that the 95% confidence limits of a result are approximately ± 40%. In Down syndrome risk screening this wide zone may be partially mitigated because the addition of extra analytes to the screen increases the spread of possible results: the likelihood ratio range for AFP only testing is 0.45 - 4.70 for an AFP range of 0.5 - 2.5 MoM. For 2 parameter testing the likelihood ratio range is 0.12 - 84.4 for an AFP and HCG range of 0.5 - 2.5 MoM.

Thus, at low levels of risk (ie. large numbers for the risk factor eg. 1:24000) there will be a wide band of uncertainty as to the ‘true’ risk but this will be of little practical significance because it is unlikely that the
uncertainty could affect the decision about whether amniocentesis should be recommended. At intermediate and high risk levels however, the size of the uncertainty zone becomes far more important because it may affect the decision about who should be offered further testing. The level at which a result can be considered intermediate depends on imprecision. For example, with a CV of 20%, the 95% confidence limits for a risk estimate are approximately n ± 40%. Thus, 2.5% of cases with a 'true' risk of 1:500 would be given a 'screen positive' result. Similarly, 2.5% of cases with a 'true risk' of 1:215 would be given a 'screen negative' result. If the CV were 40%, the cut off point where real confidence can be placed in the calculated risk becomes higher. For a risk estimate of 1:1500 with a CV of 40%, the lower 95% confidence interval is 1:300. Similarly, a positive risk must be greater than 1:165 for the 95% confidence interval to be within the 'screen positive' area. If the CV should become 50% then no risk could safely be considered 'low risk' because the 95% confidence window would extend to ± 100%, i.e. for any risk n, the confidence interval would be 1:0 to 1:2n. It is obvious therefore, that precision is vital to the effective functioning of the Down syndrome screen.

It is important to note that the simulated CVs were less than the experimentally derived CVs. This is because the simulation was based on the 'ideal' imprecision measures reported in kit inserts (AFP, 5.8%; HCG, 5.1%) whereas the experimental CV was derived from real CVs (AFP, 8.8%; HCG, 9.3%). This small increase in individual CVs is sufficient to increase the combined CV for '2-parameter' screening to levels that according to the criteria described earlier are unacceptable (> 20%). It is therefore hard to believe that the precision of routine '3-parameter' screening could ever reach a ± 20% imprecision target. It is possible that
this is the reason that some workers have shown that addition of uE₃ is at worst counter-productive in terms of increased false positives (MacDonald, 1991; Macri, 1991) and at best offers only a minimal (0.5-3%) increase in detection rate for a similar false positive rate (Nørgaard-Pedersen, 1990; Amerlite, 1991) whilst others have found no benefit to detection at all (Macri, 1990a, Spencer et al, 1992).

7.3 COMBINATION OF ASSAY AND DATING IMPRECISION (Reynolds, 1992)

It has already been shown that gestation dating errors can have a profound effect on likelihood ratio distributions. To evaluate the combined effect of assay and dating imprecision, CVs for AFP, HCG, and BPD of 6% were assumed. Since no data directly linking uE₃ and BPD was available, only 2-parameter screening was simulated. Figure 7.7 shows the effect of combining analytical imprecision with integer week based gestation dates (derived from ultrasound). This demonstrates that a very wide spread of possible likelihood ratios results from the combination of gestation age imprecision with analytical imprecision (note the LOG scale). If the BPD derived gestation dating method (or week + day gestation dating) as recommended in the conclusion to chapter 6 is used, the distribution of likelihood ratios is much narrower and does not exhibit multi-modality (Fig 7.8).

This effect of combining imprecision of assays and assay imprecision is particularly worrying because a range of likelihood ratios (0.01 - 1.0) result from the same baseline ‘target values’ using realistic CVs for laboratory and ultrasound data. In particular, they show that dating must be performed to the nearest day because weekly dating
Figure 7.7.
The effect of combining assay and dating imprecision using integer week derived MoMs.
Figure 7.8.
The effect of combining assay and dating imprecision using BPD derived MoMs.
results in a wide multimodal distribution of results. Improvements in the method of gestational age determination will significantly improve the situation (compare figure 7.7 with 7.8) but the spread of likelihood ratios is still great. Thus, we must still strive to maintain excellent levels of laboratory precision to ensure that the Down screen provides its best performance.
1) Complete assay optimisation is vital. The un-optimised assay used in the original work on uE₃ gives a completely different MoM distribution to the MoM distribution of results from the current uE₃ assay. The MoM distributions for the two AFP assays are almost identical. Since the spread of uE₃ MoMs using the new assay is much greater than using the Canick / Wald assay, this brings into question the validity of the population parameters published by Wald et al., 1988a. The standard deviation for uE₃ MoMs should be much wider than that originally published. This false narrowing of the sd for uE₃ can only result in an over-estimation of the benefits of uE₃ in Down syndrome screening with a concurrent under-estimation of the benefits of AFP and HCG. Furthermore, an un-optimised assay may alter the correlation between two analytes which could also lead to an over-estimate of its usefulness. These problems will be further considered in chapter 8.

2) Gestational age should be determined as accurately as possible and as a minimum should be quoted as week + day. Those laboratories which are not provided with precise estimates of gestational age can improve the value of the risks provided to their clinicians by reporting a range of possible risk estimates (ie. if the gestational age is given as 17 weeks, the risks calculated using medians for 16, 17 and 18 weeks could be reported). This enables the obstetric staff to evaluate more effectively whether amniocentesis should be offered.

3) Analytical imprecision should be kept to an absolute minimum. Maximum acceptable limits for imprecision have been defined as ± 10%
for most analytes and ±20% for some selected analytes. In view of the complexity of the mathematics for calculation of Down syndrome risk estimates, the greater imprecision target is reasonable. The imprecision for double analyte screening just meets this criterion but for triple analyte screening the imprecision is greater than the target. It is therefore possible that triple screening may not be valid due to an inability to achieve precision targets. Further research into whether triple testing is effective in practice is necessary.
8.1 The Effect Of Maternal Age On Screening

8.1.1 Derivation of age effects
8.1.2 Conclusions

8.2 Population Parameters

8.2.1 What values should we use?
8.2.2 Simulation-based evaluation of the most suitable parameters
8.2.3 Data-based evaluation of the most suitable parameters
8.2.4 Conclusions

8.3 Which Analytes Are Useful?

8.3.1 The second trimester
  8.3.1.1 Is Free β HCG useful?
8.3.2 The first trimester

8.4 Other Risk Calculation Algorithms

8.4.1 Assessment of atypicality
8.4.2 An examination of the discriminant function method of Nørgaard-Pedersen et al, (1990)
8.4.3 An examination of the method of Crossley et al, 1991
8.4.4 An examination of the method of Muller and Boué, 1990
8.4.5 Conclusions

8.5 The Effect Of Weight Correction On Detection

8.6 Conclusions
8.1 The Effect of Maternal Age on Screening

8.1.1 Derivation of Age Effects
(Reynolds et al, 1992b)

In most studies of Down syndrome screening, detection rates and false positive rates for the screen have been quoted as 'composite' figures: for example, 61% detection rate with 5% false positive rate using AFP, HCG and uE3 (Wald et al, 1988a); 50% detection rate with 5% false positive rate using AFP, HCG and uE3 (Nørgaard-Pedersen et al, 1990); 80% detection rate with 5% false positive rate using AFP and free β-HCG (Macri et al, 1990e). However, these rates are related to the population studied and therefore depend on the age structure of that population. Thus, they may be misleading because the Down syndrome risk factor is derived from two statistical functions, the age related risk and the likelihood ratio derived from the biochemical parameters.

Since maternal serum AFP, HCG and uE3 concentrations have been shown to be unrelated to maternal age (Wald et al, 1988a) the distribution of likelihood ratios for a population should also be unrelated to maternal age. Thus, for a given risk cut off (e.g., 1 in 300), the likelihood ratio required to modify the age related risk to indicate amniocentesis referral will vary depending on the maternal age. This has been described for screening by combination of AFP and age (Knight et al, 1988).

The physical and psychological sequelae of multiple marker serum testing have not yet been quantified but are likely to be considerable, especially for the large group of women below 35 who would not previously have been counselled about Down syndrome at all. It has
been reported that greater anxiety is experienced by women screened by AFP and subsequently referred for amniocentesis than by those undergoing amniocentesis simply because of their age (Abuelo et al, 1991). In South Wales the uptake of amniocentesis in Down screened women is approximately 85% compared to an uptake rate of approximately 35% when amniocentesis was offered on the basis of maternal age alone (Dawson et al, 1993). This also suggests that there must be a psychological effect of serum screening for Down syndrome in that serum screening results in a more positive intent to undergo a diagnostic test than when the more passive ‘age-dependent’ screening regime is used.

It takes a long time to counsel women before testing and even more time to counsel those whose result is positive. It is possible that if detection rates, false positive rates and predictive values of positive results related to the patient’s age were available they would be a valuable additional aid to counselling, since better information would be available to explain in pre-test counselling that the test does not detect all cases of Down syndrome. Furthermore, knowledge of the predictive value of a positive result may be useful when reassuring women with a positive test result that this does not definitely mean that they are carrying a Down syndrome fetus.

In the South Wales screening programs, the risk cut off used to determine which women should be offered amniocentesis is a term risk of 1:300. Since this risk (R) is derived by dividing the age related risk (r) by the likelihood ratio (LR) (i.e. \( R = \frac{r}{LR} \)), the likelihood ratio needed to change the age related risk value to the cut off value will depend on the
maternal age. Therefore using the age related risk value calculated as previously described (Cuckle et al, 1987; Reynolds and Penney, 1990), the likelihood ratio capable of causing conversion may be derived. For example: a 30 year old woman has an age related risk of a Down syndrome associated pregnancy of 1 in 909; the likelihood ratio required to convert this to the cut-off value would be 3.03 (909 / 300 = 3.03). Values for the age-related maximum likelihood ratio derived in this way are shown in Table 8.1.

To determine detection rate and false positive rate it is necessary to estimate the proportion of Down syndrome pregnancies that will be detected and the proportion of normal pregnancies that will give a false positive result. This can be achieved by a computer simulation to produce either ‘normal’ population or ‘Down syndrome’ population data as described in chapter 3. The population characteristics used were those reported in the original description of the ‘triple test’ for Down syndrome screening (Wald et al, 1988a). The simulation process was repeated 100,000 times for ‘normal’ and ‘Down syndrome’ populations for each of the three possible analytes. These simulated MoM values were used to calculate likelihood ratios for the 3 screening scenarios commonly used (AFP only, AFP + HCG, AFP + HCG + uE3) for ‘normal’ pregnancies and ‘Down syndrome’ pregnancies.

Using the simulated ‘Down syndrome’ likelihood ratios, detection rate was determined for each maternal age by finding the proportion of likelihood ratios that exceeded the age-related likelihood ratio cut-off. False positive rate was determined by the same process but using the simulated likelihood ratios from the ‘normal’ population. The simulation
Table 8.1.
Age related risk of Down syndrome (as n such that risk = 1 : n) and likelihood ratios required to modify an age risk to a risk cut off of 1:300 for women aged 16 - 44

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Risk of Down syndrome</th>
<th>Likelihood ratio to modify risk to 1 in 300</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>1572</td>
<td>5.24</td>
</tr>
<tr>
<td>18</td>
<td>1556</td>
<td>5.19</td>
</tr>
<tr>
<td>20</td>
<td>1528</td>
<td>5.09</td>
</tr>
<tr>
<td>22</td>
<td>1481</td>
<td>4.94</td>
</tr>
<tr>
<td>24</td>
<td>1404</td>
<td>4.68</td>
</tr>
<tr>
<td>26</td>
<td>1286</td>
<td>4.29</td>
</tr>
<tr>
<td>28</td>
<td>1119</td>
<td>3.73</td>
</tr>
<tr>
<td>30</td>
<td>909</td>
<td>3.03</td>
</tr>
<tr>
<td>32</td>
<td>683</td>
<td>2.28</td>
</tr>
<tr>
<td>34</td>
<td>474</td>
<td>1.58</td>
</tr>
<tr>
<td>36</td>
<td>307</td>
<td>1.02</td>
</tr>
<tr>
<td>38</td>
<td>189</td>
<td>0.63</td>
</tr>
<tr>
<td>40</td>
<td>112</td>
<td>0.37</td>
</tr>
<tr>
<td>42</td>
<td>65</td>
<td>0.22</td>
</tr>
<tr>
<td>44</td>
<td>37</td>
<td>0.12</td>
</tr>
</tbody>
</table>
technique for derivation of detection rate and false positive rate for screening by AFP alone was compared with an algebraic method that calculates an exact instead of a simulated result and was shown to give comparable results (B. Nix: personal communication). The age related detection rates and false positive rates for screening by AFP only, AFP + HCG and AFP + HCG + uE3 are shown in tables 8.2 and 8.3 respectively. Figure 8.1 is a ROC plot of this data to allow comparison of the effectiveness of the different screening combinations. It shows conclusively that using more than one analyte is necessary to achieve detection rates significantly in excess of false positive rates. It also shows that there appears to be a small increase (approx 7.4 - 7.7%) in screening efficiency when screening with 3 analytes is compared with screening with 2 analytes. This appears to contradict the predictions made in chapter 7 that suggest that addition of a third analyte will reduce screening efficiency. This is probably due to the idealised statistical model used and will be considered further in sections 8.2 and 8.3.

The only current alternative to screening by the multivariate biochemical method is to offer amniocentesis to all women. This is totally impractical and therefore the predictive value of a negative result is of little use. However, the predictive value of a positive result may be of value in counselling women whose screening result suggests referral for amniocentesis. Predictive values can be calculated using Baye’s theorem (equation 8.1). The predictive values for positive results calculated in this way relative to maternal age are shown in Table 8.4.
Table 8.2.
Age related Down syndrome detection rates (as %) for women aged 16 - 44 for screening using a 1 : 300 risk cut off.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>AFP only screening</th>
<th>AFP + HCG screening</th>
<th>AFP + HCG + uE3 screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>2.3</td>
<td>35.8</td>
<td>44.3</td>
</tr>
<tr>
<td>18</td>
<td>2.4</td>
<td>36.3</td>
<td>44.3</td>
</tr>
<tr>
<td>20</td>
<td>2.5</td>
<td>36.6</td>
<td>44.7</td>
</tr>
<tr>
<td>22</td>
<td>2.7</td>
<td>37.6</td>
<td>45.2</td>
</tr>
<tr>
<td>24</td>
<td>3.3</td>
<td>38.7</td>
<td>46.5</td>
</tr>
<tr>
<td>26</td>
<td>4.3</td>
<td>40.8</td>
<td>48.5</td>
</tr>
<tr>
<td>28</td>
<td>6.4</td>
<td>44.2</td>
<td>51.6</td>
</tr>
<tr>
<td>30</td>
<td>11.1</td>
<td>49.5</td>
<td>56.0</td>
</tr>
<tr>
<td>32</td>
<td>20.5</td>
<td>56.7</td>
<td>62.0</td>
</tr>
<tr>
<td>34</td>
<td>38.1</td>
<td>65.6</td>
<td>69.5</td>
</tr>
<tr>
<td>36</td>
<td>61.9</td>
<td>75.8</td>
<td>78.0</td>
</tr>
<tr>
<td>38</td>
<td>84.2</td>
<td>84.7</td>
<td>85.5</td>
</tr>
<tr>
<td>40</td>
<td>96.2</td>
<td>91.8</td>
<td>91.6</td>
</tr>
<tr>
<td>42</td>
<td>99.6</td>
<td>96.1</td>
<td>95.7</td>
</tr>
<tr>
<td>44</td>
<td>100.0</td>
<td>98.7</td>
<td>98.1</td>
</tr>
</tbody>
</table>
Table 8.3.
Age related false positive rates (as %) for Down syndrome screening for women aged 16 - 44 for screening using a 1 : 300 risk cut off.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>AFP only screening</th>
<th>AFP + HCG screening</th>
<th>AFP + HCG + uE₃ screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.3</td>
<td>3.0</td>
<td>3.1</td>
</tr>
<tr>
<td>18</td>
<td>0.3</td>
<td>3.0</td>
<td>3.1</td>
</tr>
<tr>
<td>20</td>
<td>0.4</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td>22</td>
<td>0.4</td>
<td>3.3</td>
<td>3.2</td>
</tr>
<tr>
<td>24</td>
<td>0.5</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>26</td>
<td>0.7</td>
<td>4.0</td>
<td>4.1</td>
</tr>
<tr>
<td>28</td>
<td>1.2</td>
<td>4.9</td>
<td>4.7</td>
</tr>
<tr>
<td>30</td>
<td>2.6</td>
<td>6.6</td>
<td>6.1</td>
</tr>
<tr>
<td>32</td>
<td>6.3</td>
<td>9.4</td>
<td>8.7</td>
</tr>
<tr>
<td>34</td>
<td>16.0</td>
<td>14.3</td>
<td>12.5</td>
</tr>
<tr>
<td>36</td>
<td>34.8</td>
<td>21.9</td>
<td>19.0</td>
</tr>
<tr>
<td>38</td>
<td>61.8</td>
<td>33.2</td>
<td>28.6</td>
</tr>
<tr>
<td>40</td>
<td>85.7</td>
<td>47.2</td>
<td>40.9</td>
</tr>
<tr>
<td>42</td>
<td>97.0</td>
<td>62.3</td>
<td>55.3</td>
</tr>
<tr>
<td>44</td>
<td>99.7</td>
<td>76.3</td>
<td>70.0</td>
</tr>
</tbody>
</table>
Figure 8.1.
ROC plot of detection and false positive rates for different screening policies.
Table 8.4.
Predictive value of a positive screening result (as %) for women aged 16 - 44 for screening using a 1 : 300 risk cut off.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>AFP only screening</th>
<th>AFP + HCG screening</th>
<th>AFP + HCG + uE₃ screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>18</td>
<td>0.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>20</td>
<td>0.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>22</td>
<td>0.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>24</td>
<td>0.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>26</td>
<td>0.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>28</td>
<td>0.5</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>32</td>
<td>0.5</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>34</td>
<td>0.5</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>36</td>
<td>0.6</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>38</td>
<td>0.7</td>
<td>1.3</td>
<td>1.6</td>
</tr>
<tr>
<td>40</td>
<td>1.0</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>42</td>
<td>1.6</td>
<td>2.3</td>
<td>2.6</td>
</tr>
<tr>
<td>44</td>
<td>2.7</td>
<td>3.4</td>
<td>3.8</td>
</tr>
</tbody>
</table>
8.1.2 Conclusions

It is well known that the incidence of Down syndrome increases with maternal age and that the serum screen is a significant advance in the prevention of the syndrome, but it is not widely recognized that the detection and false positive rates of the screen are also influenced by maternal age. This factor is extremely important because it affects the effectiveness of any Down syndrome screening program that is introduced: if the age distribution of the population of pregnant women is very young, the screen will be less effective than if the average age is older. In the US it is a common practice to offer amniocentesis to all women over 35 years and only to offer serum screening to younger women (V. Macri: personal communication). This is an important difference because detection rates in studies published by UK workers will include detection in older women who would have routinely been offered amniocentesis in the US. Failure to appreciate this could lead to an over-stating of the true overall detection rate that could falsely make it appear that screening in the US is less efficient than in Europe, simply due to a statistical property of the screening process as described above. Therefore it is vital that the lesser detection rates in younger women are
understood by all practitioners involved in Down syndrome screening.

Furthermore, it may be helpful to consider age related detection and false positive rates when counselling women prior to offering them screening. In Cardiff it is estimated to take an additional 10 minutes per woman for every pre-screen counselling session and 30 minutes to counsel those who have received a positive result (Dawson et al, 1992). The predictive values for positive results could also prove useful in counselling because they may be used to demonstrate that a positive result does not mean that the woman ‘almost certainly carries a Down syndrome child’ but in fact that there is still only a small chance depending on maternal age. For AFP screening in a 44 year old woman, only one ‘screen positive’ pregnancy in 37 will be associated with Down syndrome whilst there will be only one Down syndrome fetus per 200 ‘screen positive’ results in 16 year old women. For AFP + HCG screening these rates are one case of Down syndrome in 29 ‘screen positives’ for a 44 year old and one in 125 for a 16 year old and for AFP + HCG + uE₃ screening the rates are one in 26 for a 44 year old and one in 111 for a 16 year old. Thus, the anxiety level of the woman may be partially allayed by provision of an exact numerical risk instead of an approximation covering all women of all ages.

8.2 Population Parameters

8.2.1 What Values Should We Use?

Many different sets of population values have been published. A selection of these are shown in tables 8.5 (unaffected pregnancies) and 8.6
(Down syndrome affected pregnancies). Visual inspection shows that there are many similarities between the different sets of data. Therefore, for the purposes of this thesis only three of these sets of parameters will be evaluated: the original data published by Wald et al (1988a); the data for ultrasound dated pregnancies published by Wald et al (1992a); and the population data derived directly from the Royal Gwent Hospital and University Hospital of Wales screening programs.

8.2.2 Simulation-Based Evaluation of the Most Suitable Parameters

To determine the screening characteristics resulting from each set of population parameters, it is necessary to determine the proportions of unaffected and Down syndrome pregnancies that have the same risk estimate. Simulation of maternal age is unnecessary because the distribution of maternal ages is defined and for the simulation to be effective it must be large enough for the age distribution for any particular combination of parameters to match the defined distribution. Therefore, by not simulating the age distribution the total number of data sets in the simulation may be reduced allowing a fair comparison of the performance of the different population data sets although not necessarily a good estimate of detection rates accruing from their use.

Using the computer subroutines in appendix 1 (described in section 3.3) and the population parameters described in section 8.2.1, 16000 sets of AFP, HCG and uE₃ results for unaffected and Down syndrome pregnancies were simulated. To ensure comparability, the same random seeds (1546, 15782, 751) were used for all simulations. The analyte results were used to derive likelihood ratios for 'double' (AFP + HCG) and 'triple'
Table 8.5.
Reported population parameters for unaffected pregnancies. Note: uE₃ results marked * were not LOG transformed. Values in brackets are the population standard deviations.

<table>
<thead>
<tr>
<th>Correlations</th>
<th>AFP</th>
<th>HCG</th>
<th>uE₃</th>
<th>A/H</th>
<th>A/u</th>
<th>H/u</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald et al, 1988a (n = 365)</td>
<td>0.0073 (0.2024)</td>
<td>0.0 (0.2342)</td>
<td>*1.0 (0.27)</td>
<td>0.05</td>
<td>*0.13</td>
<td>*-0.08</td>
</tr>
<tr>
<td>Heyl et al, 1990 (n = 85)</td>
<td>0.006 (0.113)</td>
<td>-0.045 (0.242)</td>
<td>*1.03 (0.26)</td>
<td>0.16</td>
<td>*0.16</td>
<td>*0.01</td>
</tr>
<tr>
<td>Nørgaard-P et al, 1990 (n = 328)</td>
<td>0.01 (0.152)</td>
<td>0.009 (0.248)</td>
<td>0.003 (0.118)</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crossley et al, 1991 (n = 410)</td>
<td></td>
<td>0.0043 (0.2499)</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zeitune et al, 1991 (n = 113000)</td>
<td></td>
<td>-0.0044 (0.1668)</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spencer, 1991 (n = 145)</td>
<td>0.0031 (0.156)</td>
<td>0.0042 (0.2361)</td>
<td>0.0028 (0.1572)</td>
<td>0.2802</td>
<td>0.1886</td>
<td>-0.1916</td>
</tr>
<tr>
<td>Spencer et al, 1992 (n = 2862)</td>
<td>-0.0092 (0.1931)</td>
<td>0.0132 (0.241)</td>
<td>-0.016 (0.148)</td>
<td>0.152</td>
<td>0.305</td>
<td>-0.119</td>
</tr>
<tr>
<td>Ryall et al, 1992 (n = 171)</td>
<td>0.0 (0.1542)</td>
<td>0.0 (0.2879)</td>
<td>0.0 (0.1472)</td>
<td>0.164</td>
<td>0.317</td>
<td>-0.216</td>
</tr>
<tr>
<td>Wald et al, 1992a (n = 20507)</td>
<td>0.0 (0.191)</td>
<td>0.0 (0.2401)</td>
<td>0.0 (0.1184)</td>
<td>0.1759</td>
<td>0.104</td>
<td>-0.0476</td>
</tr>
<tr>
<td>Crossley et al, 1993 (n = 390)</td>
<td></td>
<td>0.0 (0.13)</td>
<td>0.25</td>
<td>-0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Royal Gwent / UHW (n = 5082 (AFP / HCG); n = 536 uE₃)</td>
<td>0.0 (0.177)</td>
<td>0.0 (0.239)</td>
<td>0.0 (0.186)</td>
<td>0.183</td>
<td>0.130</td>
<td>-0.08</td>
</tr>
</tbody>
</table>
Table 8.6.
Reported population parameters for Down syndrome affected pregnancies. Note: uE₃ results marked * were not LOG transformed. Values in brackets are the population standard deviations.

<table>
<thead>
<tr>
<th>Study</th>
<th>Correlations</th>
<th>AFP</th>
<th>HCG</th>
<th>uE₃</th>
<th>A/H</th>
<th>A/u</th>
<th>H/u</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald et al, 1988a (n = 77)</td>
<td></td>
<td>-0.1427 (0.2052)</td>
<td>0.3096 (0.2588)</td>
<td>*0.73 (0.26)</td>
<td>0.14</td>
<td>*0.14</td>
<td>*-0.25</td>
</tr>
<tr>
<td>Heyl et al, 1990 (n = 16)</td>
<td></td>
<td>-0.112 (0.162)</td>
<td>0.29 (0.283)</td>
<td>*0.79 (0.33)</td>
<td>0.46</td>
<td>*0.08</td>
<td>*0.46</td>
</tr>
<tr>
<td>Nørgaard-P et al, 1990 (n = 42)</td>
<td></td>
<td>-0.154 (0.198)</td>
<td>0.195 (0.317)</td>
<td>-0.129 (0.202)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crossley et al, 1991 (n = 49)</td>
<td></td>
<td></td>
<td></td>
<td>0.3385 (0.3127)</td>
<td></td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Zeitune et al, 1991 (n = 114)</td>
<td></td>
<td>-0.1427 (0.1626)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spencer, 1991 (n = 29)</td>
<td></td>
<td>-0.13 (0.2028)</td>
<td>0.2881 (0.2257)</td>
<td>-0.1879 (0.211)</td>
<td>-0.1192</td>
<td>0.6235</td>
<td>-0.3267</td>
</tr>
<tr>
<td>Spencer et al, 1992 (n = 90)</td>
<td></td>
<td>-0.1413 (0.2013)</td>
<td>0.3282 (0.2825)</td>
<td>-0.1601 (0.2101)</td>
<td>-0.188</td>
<td>0.374</td>
<td>-0.293</td>
</tr>
<tr>
<td>Ryall et al, 1992 (n = 57)</td>
<td></td>
<td>-0.1281 (0.2015)</td>
<td>0.3262 (0.2462)</td>
<td>-0.1533 (0.1785)</td>
<td>0.049</td>
<td>0.478</td>
<td>-0.22</td>
</tr>
<tr>
<td>Wald et al, 1992a (n = 77)</td>
<td></td>
<td>-0.1427 (0.194)</td>
<td>0.3023 (0.2665)</td>
<td>-0.1411 (0.1741)</td>
<td>0.2392</td>
<td>0.0521</td>
<td>-0.2106</td>
</tr>
<tr>
<td>Crossley et al, 1993 (n = 49)</td>
<td></td>
<td></td>
<td></td>
<td>-0.102 (0.156)</td>
<td></td>
<td>0.44</td>
<td>0.11</td>
</tr>
<tr>
<td>Royal Gwent / UHW (n = 52)</td>
<td></td>
<td>-0.142 (0.174)</td>
<td>0.318 (0.269)</td>
<td>-0.146 (0.195)</td>
<td>0.295</td>
<td>0.047</td>
<td>-0.189</td>
</tr>
</tbody>
</table>
(AFP + HCG + uE₃) screening. These likelihood ratios were imported into a spreadsheet (Microsoft Excel) and sorted into ascending order. The Down syndrome population of likelihood ratios was examined to determine the proportions corresponding to the 0.5 - 10.0 percentiles of the unaffected distribution (at 0.5 percentile intervals). The values for 'double' testing are shown in table 8.7 and for triple testing in table 8.8.

In table 8.7 it is evident that the Gwent derived parameters for 'double' testing appear theoretically superior to the parameters published by Wald et al. In table 8.8, the Wald et al, 1992a parameters appear theoretically superior for the purposes of 'triple' screening. These tables may also be used to determine the marginal benefit of adding uE₃. It is generally taken that 5% is an acceptable amniocentesis rate and therefore the marginal benefit of uE₃ will be considered at the 5th percentile. For the Wald et al, (1988a) parameters, the benefit is +8% (similar to the benefit estimated in section 8.1.1); for the Wald et al, (1992a) parameters, the benefit is +11.4%; and for the Gwent parameters, the benefit is +3.1%. The lesser benefit of adding uE₃ to the Gwent parameters is interesting particularly because the Gwent 'double' test parameters have exactly the same efficiency at the 5th percentile as the Wald et al, (1988a) triple test parameters. The Wald et al, (1992a) triple test parameters still however suggest a +6.9% benefit of adding uE₃ over the Gwent 'double' test parameters.

To confirm the veracity of this method of comparing the 3 sets of screening parameters, it is useful to examine the graph in figure 8.1 and the age related figures from which it is derived. These figures were derived using the parameters of Wald et al, (1988a) and for a 28 year old
Table 8.7.
Proportions of simulated Down syndrome population having the same likelihood ratio as the unaffected population for a 'double' test.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentile</td>
<td>0.5</td>
<td>1.0</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>17.1</td>
<td>23.2</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>15.4</td>
<td>21.1</td>
<td>28.4</td>
</tr>
<tr>
<td></td>
<td>19.6</td>
<td>26.5</td>
<td>31.5</td>
</tr>
<tr>
<td></td>
<td>25.9</td>
<td>28.4</td>
<td>31.5</td>
</tr>
<tr>
<td></td>
<td>29.2</td>
<td>32.0</td>
<td>35.4</td>
</tr>
<tr>
<td></td>
<td>29.2</td>
<td>32.0</td>
<td>35.4</td>
</tr>
<tr>
<td></td>
<td>32.0</td>
<td>34.9</td>
<td>38.4</td>
</tr>
<tr>
<td></td>
<td>34.5</td>
<td>37.2</td>
<td>41.6</td>
</tr>
<tr>
<td></td>
<td>39.9</td>
<td>44.1</td>
<td>46.9</td>
</tr>
<tr>
<td></td>
<td>42.6</td>
<td>46.8</td>
<td>51.3</td>
</tr>
<tr>
<td></td>
<td>46.9</td>
<td>48.7</td>
<td>52.9</td>
</tr>
<tr>
<td></td>
<td>50.5</td>
<td>50.5</td>
<td>54.5</td>
</tr>
<tr>
<td></td>
<td>52.4</td>
<td>54.5</td>
<td>56.3</td>
</tr>
<tr>
<td></td>
<td>53.8</td>
<td>57.8</td>
<td>57.8</td>
</tr>
<tr>
<td></td>
<td>55.1</td>
<td>59.3</td>
<td>59.3</td>
</tr>
<tr>
<td></td>
<td>56.4</td>
<td>60.5</td>
<td>60.5</td>
</tr>
<tr>
<td></td>
<td>58.8</td>
<td>61.9</td>
<td>61.9</td>
</tr>
<tr>
<td></td>
<td>60.2</td>
<td>63.0</td>
<td>63.0</td>
</tr>
<tr>
<td></td>
<td>61.3</td>
<td>64.1</td>
<td>64.1</td>
</tr>
<tr>
<td></td>
<td>65.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8.8.
Proportions of simulated Down syndrome population having the same likelihood ratio as the unaffected population for a 'triple' test.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentile</td>
<td>20.2</td>
<td>30.3</td>
<td>22.4</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>28.3</td>
<td>37.8</td>
<td>29.0</td>
</tr>
<tr>
<td>1.5</td>
<td>33.5</td>
<td>42.2</td>
<td>34.9</td>
</tr>
<tr>
<td>2.0</td>
<td>37.3</td>
<td>45.6</td>
<td>38.6</td>
</tr>
<tr>
<td>2.5</td>
<td>40.6</td>
<td>48.2</td>
<td>41.6</td>
</tr>
<tr>
<td>3.0</td>
<td>42.5</td>
<td>50.2</td>
<td>45.0</td>
</tr>
<tr>
<td>3.5</td>
<td>44.8</td>
<td>52.7</td>
<td>47.8</td>
</tr>
<tr>
<td>4.0</td>
<td>47.4</td>
<td>54.6</td>
<td>50.0</td>
</tr>
<tr>
<td>4.5</td>
<td>49.0</td>
<td>56.7</td>
<td>52.4</td>
</tr>
<tr>
<td>5.0</td>
<td>51.3</td>
<td>58.2</td>
<td>54.4</td>
</tr>
<tr>
<td>5.5</td>
<td>53.1</td>
<td>59.9</td>
<td>56.1</td>
</tr>
<tr>
<td>6.0</td>
<td>54.8</td>
<td>61.5</td>
<td>57.6</td>
</tr>
<tr>
<td>6.5</td>
<td>56.3</td>
<td>62.5</td>
<td>59.2</td>
</tr>
<tr>
<td>7.0</td>
<td>57.9</td>
<td>63.7</td>
<td>60.6</td>
</tr>
<tr>
<td>7.5</td>
<td>58.8</td>
<td>64.7</td>
<td>61.8</td>
</tr>
<tr>
<td>8.0</td>
<td>60.1</td>
<td>65.6</td>
<td>63.4</td>
</tr>
<tr>
<td>8.5</td>
<td>61.5</td>
<td>66.6</td>
<td>64.5</td>
</tr>
<tr>
<td>9.0</td>
<td>62.7</td>
<td>67.6</td>
<td>65.7</td>
</tr>
<tr>
<td>9.5</td>
<td>63.8</td>
<td>68.6</td>
<td>67.0</td>
</tr>
<tr>
<td>10.0</td>
<td>65.1</td>
<td>69.5</td>
<td>67.9</td>
</tr>
</tbody>
</table>
woman, the estimated detection and false positive rates are 44.2% detection for a 4.9% false positive rate for 'double' testing and 51.6% detection for a 4.7% false positive rate for 'triple' testing. These figures are very close to the 5th percentile values in tables 8.7 and 8.8. This means that the values in the tables can be viewed as fair theoretical estimates of detection for different false positive rates and allows comparison with the estimates for detection previously published.

The Wald et al, (1988a) report presented a table of false positive rate compared to fixed levels of detection which means that it is difficult to compare them visually with the results derived here. To simplify comparison, the two sets of screening figures are presented as a ROC plot (figure 8.2). This suggests that the original estimates of detection were over-optimistic when compared to the estimates from the simulation (approximately +12% on detection for 'double' testing and +10% for 'triple' testing). The estimates of detection for scan dated pregnancies in the Wald et al, (1992a) report are compared with the Wald et al, (1992a) parameter derived estimates of detection in figure 8.3. This also shows an apparent +12% excess expectation for 'double testing and +9% for triple testing. However, this may simply be due different methods of evaluating detection: The method described here is independent of maternal age whereas Wald integrated his results against the age distribution of pregnant women in the UK in 1981-5. Since detection is greater in older women (as described above) this could explain the differences.
Figure 8.2.
ROC plot comparing data derived in section 8.2.2 and from Wald et al, 1988a.
Figure 8.3.
ROC plot comparing data derived in section 8.2.2 and from Wald et al, 1992a.
8.2.3 Data-Based Evaluation of the Most Suitable Parameters

To evaluate the performance of the different screening parameters using real data, a set of results from unaffected and Down syndrome affected pregnancies was identified. There were 536 unaffected results and 52 Down syndrome results. The unaffected results and 27 of the Down syndrome results came from the UHW and RGH screening programs and the remainder of the Down syndrome results came from Oldchurch Hospital, Romford (Kevin Spencer: Personal communication). Likelihood ratios for each set of results were derived using the Wald et al (1988a and 1992a) population parameters as described in section 8.2.2. This was not done using the Gwent parameters because these were derived from the same data set and it is obvious that testing a data set against itself could lead to an over-estimate of the benefits of that dat sets' population characteristics.

Due to the small size of the data set, the maternal age corresponding to each likelihood ratio was not used to derive a Down syndrome risk as this would lead to incremental jumping in the detection and false positive rates. To compensate for this and estimate the overall population effects of screening, the likelihood ratios were numerically integrated with the age distribution of pregnant women in the UK, as first described by Wald et al (1988a). The detection and false positive rates for different risk cut-offs are shown in table 8.9. Since it is difficult to compare these visually they are also shown as a ROC plot in figure 8.4.

From the figure it is immediately apparent that three of the lines are so similar as to be almost indistinguishable and that the fourth line relates
Table 8.9
Detection and false positive rates for 'double' and 'triple' testing using population parameters from Wald et al, 1988a and 1992a.

<table>
<thead>
<tr>
<th>Risk</th>
<th>False positive rates (%)</th>
<th>Detection rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wald et al, 1988a</td>
<td>Wald et al, 1992a</td>
</tr>
<tr>
<td></td>
<td>'double'</td>
<td>'triple'</td>
</tr>
<tr>
<td>1: 100</td>
<td>1.67</td>
<td>2.07</td>
</tr>
<tr>
<td>1: 150</td>
<td>2.70</td>
<td>3.18</td>
</tr>
<tr>
<td>1: 200</td>
<td>3.88</td>
<td>4.52</td>
</tr>
<tr>
<td>1: 250</td>
<td>4.88</td>
<td>5.69</td>
</tr>
<tr>
<td>1: 300</td>
<td>5.81</td>
<td>6.80</td>
</tr>
</tbody>
</table>

Wald et al, 1988a
Wald et al, 1992a
Figure 8.4.
ROC plot comparing 'double' and 'triple' testing using Wald et al, 1988a and 1992a population parameters.
Open markers represent screening using Wald et al, 1988a parameters, Closed markers represent screening using Wald et al, 1992a parameters. Circles represent 'double' testing, squares represent 'triple testing.'
to a screening test that is significantly less efficient (approximately -6% detection for a constant false positive rate). From the theoretical evaluation in section 8.2.2, it is apparent that the difference between 'double' testing using the Wald 1988a and 1992a parameters is approximately +3.5% in favour of the 1992a parameters. It is therefore not particularly surprising that this small difference is not seen in the 'real' data. It is however much more surprising that the addition of uE₃ adds nothing when the 1988a parameters are used (actual benefit, 0%, expected benefit +8%) and causes a loss of detection when the 1992a parameters are used (actual benefit, -6%, expected benefit +11.4%). This suggests that although in theory addition of extra analytes adds detection, in practice other factors, possibly including the imprecision problem discussed in chapter 7, section 7.2 mean that addition of extra analytes may not be useful.

This demonstration of the lack of benefit of adding uE₃ is supported by the clinical trial performed by Amerlite Diagnostics Ltd., where using likelihood ratios derived from trials data (2765 unaffected pregnancies, 126 Down syndrome pregnancies), the detection rate for a 5% false positive rate was 55.4% for 'double' testing and 55.0% for 'triple' testing (Amerlite Diagnostics, 1991). Interestingly, after multivariate Gaussian simulation of the trials data the detection rates (± 95% confidence limit) for a 5% false positive rate were 57.3 ± 5.8% for 'double' testing and 57.8 ± 6.4% for 'triple' testing. The close similarity of these two results is further evidence that it may be unnecessary to add uE₃ to the 'double' test.

Before uE₃ can be completely rejected however, it would be helpful to determine the reason why the 1992a parameters cause a loss of
detection. The starting point for the investigation of this problem is the population parameters themselves: When the Wald 1992a and Gwent parameters are compared it can be seen that the figures for affected pregnancies are quite similar with the only difference being a slightly smaller sd for AFP and greater sd for uE₃ in the Gwent data; When the unaffected pregnancy data is examined the AFP and HCG values are also closely similar but the outstanding feature is that the sd for uE₃ is significantly smaller in the Wald 1992a data (0.1184 compared to 0.186). This offers a good explanation for the poorer performance of the triple test using the Wald 1992a parameters: Since the 'real' population sd was greater than the 'model' population sd a greater proportion of unaffected pregnancies than the model predicts appear to be further away from the centre of the 'model' distribution (figure 8.5). Obviously, half of these results will be on the Down syndrome side of the distribution and thus the model's prediction that they are further away from the unaffected mean causes an increased number of false positive diagnoses. Since we are considering detection for a constant false positive rate, any increase in false positives will cause an apparent decrease in detection. This is an important lesson because it shows that predictions of screening based on an 'idealised' statistically simulated population model can be completely wrong when 'real' population data is applied.

8.2.4 Conclusions

1) The correct population parameters are essential: Use of incorrect parameters can result in a significant loss of detection.

2) Computer simulations of population parameters have been used to
Unaffected population members who appear to be more like members of the affected population. Result is more false positives.

Figure 8.5.
Diagram to show why too small a population 'model' sd causes a loss of detection. For clarity the separation between affected and unaffected populations has been accentuated. See figure 3.2.
determine detection and false positive rates. If this method of
determination has been used the predicted detection and false positive
rates should be viewed with some caution unless they are confirmed by
testing against a separate data set.

3) Simulation using the Gwent parameters suggests that adding a third
analyte (uE₃) only produces a benefit of +3.5% in comparison with other
investigators who have reported benefits up to +11.4%. Data-based
analysis of false positive and detection rates suggest that these higher
rates may be over-optimistic and that there is possibly only little benefit
of adding a third analyte. Since a separate data set was not available to
test the Gwent parameters no firm conclusion about their utility may be
drawn.

8.3 Which Analytes Are Useful?

8.3.1 The Second Trimester

In sections 7.2 and 8.2 evidence has been presented that suggests
that unless analytical imprecision can be markedly reduced, Down
syndrome risk estimates should only be derived using the results for
maternal age and two biochemical analytes. In section 8.2 it was assumed
that the basic screening combination should be AFP + HCG and that the
third analyte that should only be used if it is proved that it adds
significantly to detection. The assumption that AFP and HCG should be
used in a 'double' test is based on established clinical practice: in the
second trimester, AFP is already assayed in a large number of
laboratories for the purposes of neural tube defect screening. It is well
recognised that the combination of AFP and uE₃ is not efficient (Wald et
al, 1988a; Amerlite Diagnostics, 1991) and therefore the second analyte must be HCG. This was the case in South Wales when Down syndrome screening was introduced. Furthermore, the combination of uE₃ with HCG as a 'double' test has been shown to be less efficient than AFP + HCG (Amerlite Diagnostics, 1991). Therefore for 'double' testing in the second trimester there is only one acceptable combination: AFP + HCG. The data above has all been in relation to AFP + total HCG but reports have been published that suggest that the free β subunit of HCG is superior (Macri et al, 1990e: Ryall et al, 1992: Spencer, 1991: Spencer, 1992: Spencer et al, 1992).

8.3.1.1 Is Free β HCG Useful?
(Stone et al, In press)

To evaluate the proposal that free β HCG is superior to total HCG, samples from 21 Down syndrome and 180 unaffected pregnancies were analysed for free β-HCG using the CIS (UK) Ltd., free β-HCG assay kit (section 3.1.4). The Down syndrome samples were collected prospectively at the UHW and had been stored at -20°C. The maternal ages, AFP, total HCG, free β-HCG MoMs for the Down syndrome samples and the risks derived for the combination of AFP + total HCG (using the Wald et al, 1988a) parameters and for the combination of AFP + free β-HCG (population parameters for free β-HCG, K. Spencer, personal communication) are shown in table 8.10.

The results are divided into three groups based on a 1:300 risk cut off: those that would have been detected with either protocol, those that would have been missed by both protocols and those that would be
Table 8.10
Comparison of AFP + total HCG screening (Risk 1) with AFP + free-β HCG screening (Risk 2). The values in the risk columns are n such that risk = 1 in n.

<table>
<thead>
<tr>
<th>Age @ EDD (MoM)</th>
<th>AFP MoM</th>
<th>HCG MoM</th>
<th>Free-β HCG MoM</th>
<th>Risk 1 n</th>
<th>Risk 2 n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects identified by both protocols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.5</td>
<td>0.46</td>
<td>5.35</td>
<td>3.64</td>
<td>49</td>
<td>70</td>
</tr>
<tr>
<td>34.4</td>
<td>1.11</td>
<td>3.0</td>
<td>3.65</td>
<td>100</td>
<td>81</td>
</tr>
<tr>
<td>31.6</td>
<td>0.45</td>
<td>1.45</td>
<td>2.74</td>
<td>300</td>
<td>84</td>
</tr>
<tr>
<td>19.0</td>
<td>0.81</td>
<td>7.95</td>
<td>5.94</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>36.2</td>
<td>0.52</td>
<td>1.3</td>
<td>1.35</td>
<td>260</td>
<td>190</td>
</tr>
<tr>
<td>40.0</td>
<td>0.68</td>
<td>2.0</td>
<td>1.57</td>
<td>36</td>
<td>88</td>
</tr>
<tr>
<td>40.0</td>
<td>0.94</td>
<td>3.0</td>
<td>2.95</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>35.3</td>
<td>0.65</td>
<td>3.55</td>
<td>5.59</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>34.9</td>
<td>0.97</td>
<td>3.9</td>
<td>5.02</td>
<td>52</td>
<td>21</td>
</tr>
<tr>
<td>38.7</td>
<td>0.6</td>
<td>2.8</td>
<td>2.73</td>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td>39.1</td>
<td>0.4</td>
<td>5.53</td>
<td>8.3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>39.4</td>
<td>0.75</td>
<td>1.1</td>
<td>1.43</td>
<td>210</td>
<td>140</td>
</tr>
<tr>
<td>32.6</td>
<td>0.5</td>
<td>2.0</td>
<td>2.5</td>
<td>149</td>
<td>65</td>
</tr>
<tr>
<td>30.4</td>
<td>0.16</td>
<td>0.55</td>
<td>1.11</td>
<td>206</td>
<td>84</td>
</tr>
</tbody>
</table>

| Subjects identified by only one protocol |
| 23.9            | 1.03    | 2.37    | 5.1            | 620     | 76      |
| 29.5            | 0.94    | 2.8     | 1.93           | 280     | 830     |

| Subjects not identified by either protocol |
| 29.2            | 0.79    | 1.41    | 0.9            | 980     | 2700    |
| 35.3            | 0.65    | 0.8     | 0.9            | 1200    | 670     |
| 33.1            | 1.1     | 1.75    | 2.13           | 660     | 490     |
| 26.6            | 1.2     | 2.0     | 1.89           | 1100    | 1800    |
| 18.9            | 0.63    | 1.32    | 1.27           | 1400    | 1600    |
detected by only one protocol. It can be seen that using either AFP + HCG or AFP + free-β HCG and a 1:300 cut off, 15 of the Down syndrome cases would have been detected (detection rate 71%). It should be noted that there was one discordant pair where total HCG screening identified one case that free-β HCG screening missed and vice-versa. However, the false positive rates for the two methods are significant: for the AFP + total HCG combination, 5% of the 180 unaffected samples gave a 'high risk' result compared to 7.5% for the AFP + free-β HCG combination. This compares with the initial screen positive rate in the South Wales screening program which is 3.5% (296 screen positive results in 8414 tests) (Dawson et al, 1992).

This study is very small and can therefore only be used as a rough guide to the performance of free-β HCG in Down syndrome screening but it appears that using the UHW data, free-β HCG is less effective than total HCG because it increases the false positive rate. The published literature on free-β HCG is still quite small and argument about the benefit of using it is continuing: Using the same population parameters (those of Spencer et al, 1992), Cuckle and Lilford (1992, 1993) claim that free-β offers a 8-10% increase in detection for a constant false positive rate whilst Wald and Hackshaw, (1993) claim that there is only a 5-6% increase. This debate will continue until more data has been published. The effectiveness of other markers is also to be determined. Some of these markers have already been briefly discussed in chapter 2 and will not be further considered here.
8.3.2 The first trimester

Screening in the second trimester of pregnancy is a race against time: blood samples are collected at approximately 16 weeks of gestation; if a 'high risk' result is found, an amniocentesis must be arranged and the cells from the amniotic fluid collection must be cultured before a cytogenetic diagnosis can be made. The gestation age limit beyond which abortion is not allowed in the UK (except in the instance of a severely abnormal fetus) is 24 weeks. An argument may be made that Down syndrome is a severe abnormality but notwithstanding this, a late abortion is an extremely unpleasant procedure (A. Dawson, personal communication) and should be avoided if possible. Therefore, the possibility that Down syndrome screening could be moved to the second trimester has been investigated. Markers of potential use in the first trimester will be considered in the order AFP, HCG (including free-β HCG), and uE₃.


HCG in the first trimester has also been examined by many investigators and it has been shown that there is less difference between Down syndrome and unaffected pregnancies in the first trimester (Brock et al, 1990: Johnson et al, 1991: Kratzer et al, 1991: Van Lith, 1992). The
difference in levels of free-β HCG however are maintained (Nebbiolo et al, 1990: Spencer et al, 1992b). It would therefore appear that in the first trimester free-β HCG is the better marker.

Unconjugated estriol in the first trimester has been examined by a small number of investigators (Cuckle et al, 1988: Brock et al, 1990: Crandall et al, 1991) but was assayed using an Amerlex-M assay that was modified in a similar fashion to the Canick / Wald modification (as described in section 3.1.3.5) and therefore although the results suggest that uE₃ is also lower in the first trimester (approx 0.75 MoM), these results are suspect because the assays were not designed for measuring uE₃ at the low concentrations in the first trimester. Therefore the problems described in section 7.1 apply equally to the data on first trimester uE₃ and caution must be taken when considering whether uE₃ is useful for screening in the first trimester.

Other markers that have been examined in the first trimester are Ca125 (Van Lith, 1991), pregnancy specific β₁-glycoprotein (SP₁) (Brock et al, 1990), placental alkaline phosphatase (Brock et al, 1990), PAPP-A (Wald et al, 1992c), free α-HCG (Kratzer et al, 1991) and progesterone (Kratzer et al, 1991) but published data is too limited for any conclusions to be made about whether any of these markers may be useful. Therefore at present the best combination of analytes for first trimester screening appears to be AFP + free-β HCG although as more data emerges this may change.
8.4 Other Risk Calculation Algorithms

8.4.1 Assessment of Atypicaility  
(Wright et al, In press)

When Down syndrome risk estimates are calculated they are precisely that: an estimate of the risk of Down syndrome. Thus, if a condition that causes changes in the screening analytes that are different to those caused by Down syndrome is present, the final risk estimate may be very low which could result in a false sense of security. In some such conditions the analyte levels may be sufficiently abnormal to be identified by the biochemistry or obstetric staff but this may not always be the case. To get around this problem it is possible to add specific screens that look for other conditions such as trisomy 18 and trisomy 13. However, these conditions are significantly rarer than Down syndrome and it is therefore difficult to collect sufficient data to derive the population parameters needed to perform specific screens. Even if this data were to be available, there would still be the possibility that other conditions would be missed. Thus a non-specific screen that detects those cases where the results are significantly atypical of normality despite having a low risk of Down syndrome would be useful.

Equation 8.2

<table>
<thead>
<tr>
<th>Case</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate</td>
<td>( d = \frac{(x - \mu)^2}{\sigma^2} )</td>
</tr>
<tr>
<td>Bivariate</td>
<td>( d = \left( \frac{x-h_x}{\sigma_x} \right)^2 + \left( \frac{y-h_y}{\sigma_y} \right)^2 - 2\rho \left( \frac{x-h_x}{\sigma_x} \right) \left( \frac{y-h_y}{\sigma_y} \right) )</td>
</tr>
<tr>
<td>Trivariate</td>
<td>( d = (x - \mu) \cdot \nabla \cdot (x - \mu) )</td>
</tr>
</tbody>
</table>
The non-specific screen described here determines whether the pattern of results is significantly atypical of normality. Essentially, the screen uses the Mahalanobis distance (d) (Gower, 1985). This is already calculated in the Gaussian distribution functions used for the calculations in the likelihood ratio method (equations 3.6, 3.7 and 3.8 as shown in equation 8.2).

The Mahalanobis distance is distributed according to the $\chi^2$ distribution and therefore if the computed distance exceeds the upper $(1 - \alpha)$ 100% quantile of the $\chi^2$ distribution with $v = p$ degrees of freedom (where $p =$ the number of analytes used in the calculation). Thus for a bivariate situation, the cut offs for 95%, 98% and 99% are 5.991, 7.824, and 9.210 respectively.

The effect of using the Mahalanobis distance to distinguish atypical results was evaluated by calculating Down syndrome risks and Mahalanobis distances for the results of 2000 normal pregnancies and 37 abnormal (not trisomy 21) pregnancies from the literature (Bogart et al, 1987; Staples et al, 1991, Johnson et al, 1991). The population parameters used to calculate these values were from Wald et al, 1988a. The results for the abnormal pregnancies are shown in table 8.11 and the result interpretation algorithm is shown in figure 8.6. A breakdown of the effect of using the interpretation algorithm is shown in table 8.12. It is obvious that the amended interpretation algorithm is very effective at deciding when there is trisomy 18 or some other chromosomal anomalies but poor at detecting trisomy 13. The reason why trisomy 18 is better discriminated may be seen more clearly in figure 8.7. The trisomy 13 cases are much more typical of normality than the trisomy 18 cases. It is
Table 8.11. Down syndrome risks and Mahalanobis distances for 37 non-Down syndrome pregnancies.

<table>
<thead>
<tr>
<th>Abnormality</th>
<th>$\log_e(\text{MoM})$ AFP</th>
<th>$\log_e(\text{MoM})$ HCG</th>
<th>Down's Risk</th>
<th>Mahalanobis distance Normal</th>
<th>Mahalanobis distance Down's</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trisomy 13</td>
<td>1.279</td>
<td>1.669</td>
<td>50</td>
<td>14.33</td>
<td>9.98</td>
</tr>
<tr>
<td>Trisomy 13</td>
<td>0.029</td>
<td>0.425</td>
<td>510</td>
<td>0.67</td>
<td>0.78</td>
</tr>
<tr>
<td>Trisomy 13</td>
<td>-1.386</td>
<td>-0.994</td>
<td>790</td>
<td>8.86</td>
<td>9.85</td>
</tr>
<tr>
<td>Trisomy 13</td>
<td>-0.693</td>
<td>-0.799</td>
<td>1130</td>
<td>5.29</td>
<td>7.00</td>
</tr>
<tr>
<td>Trisomy 13</td>
<td>-0.693</td>
<td>-0.431</td>
<td>550</td>
<td>4.08</td>
<td>4.36</td>
</tr>
<tr>
<td>Trisomy 13</td>
<td>-0.117</td>
<td>-0.288</td>
<td>1870</td>
<td>0.35</td>
<td>3.07</td>
</tr>
<tr>
<td>Trisomy 13</td>
<td>0.513</td>
<td>-0.248</td>
<td>4770</td>
<td>2.37</td>
<td>6.96</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.041</td>
<td>-2.303</td>
<td>36300</td>
<td>18.53</td>
<td>27.18</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0.010</td>
<td>-1.897</td>
<td>27600</td>
<td>12.65</td>
<td>20.76</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.400</td>
<td>-1.661</td>
<td>9330</td>
<td>10.03</td>
<td>15.96</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0.000</td>
<td>-1.561</td>
<td>1890</td>
<td>8.52</td>
<td>15.84</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0.000</td>
<td>-1.204</td>
<td>11600</td>
<td>5.05</td>
<td>11.42</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0.223</td>
<td>-1.139</td>
<td>15100</td>
<td>5.23</td>
<td>12.13</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0.285</td>
<td>-0.478</td>
<td>5550</td>
<td>1.56</td>
<td>6.46</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0.916</td>
<td>-0.163</td>
<td>5450</td>
<td>6.75</td>
<td>11.61</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0.916</td>
<td>-0.051</td>
<td>4380</td>
<td>6.53</td>
<td>10.95</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.329</td>
<td>-0.329</td>
<td>1860</td>
<td>1.10</td>
<td>3.00</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.371</td>
<td>-1.609</td>
<td>36700</td>
<td>9.35</td>
<td>15.29</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.634</td>
<td>-1.470</td>
<td>7020</td>
<td>9.48</td>
<td>13.70</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.755</td>
<td>-2.659</td>
<td>30100</td>
<td>26.62</td>
<td>32.62</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.198</td>
<td>-0.073</td>
<td>2330</td>
<td>0.33</td>
<td>1.78</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.416</td>
<td>-1.966</td>
<td>47800</td>
<td>13.82</td>
<td>20.37</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.058</td>
<td>-0.580</td>
<td>12300</td>
<td>1.21</td>
<td>5.75</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.580</td>
<td>-1.609</td>
<td>22400</td>
<td>10.46</td>
<td>15.36</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.673</td>
<td>-0.799</td>
<td>2130</td>
<td>5.09</td>
<td>6.93</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.344</td>
<td>-0.030</td>
<td>5850</td>
<td>0.89</td>
<td>41.7</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.248</td>
<td>-0.616</td>
<td>8310</td>
<td>1.60</td>
<td>5.02</td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>-0.713</td>
<td>-2.526</td>
<td>34380</td>
<td>23.98</td>
<td>30.02</td>
</tr>
<tr>
<td>45, XO</td>
<td>0.565</td>
<td>1.601</td>
<td>50</td>
<td>10.40</td>
<td>5.73</td>
</tr>
<tr>
<td>47, XXY</td>
<td>0.378</td>
<td>0.571</td>
<td>650</td>
<td>1.97</td>
<td>2.57</td>
</tr>
<tr>
<td>69, XXX</td>
<td>-1.109</td>
<td>-2.303</td>
<td>4490</td>
<td>22.36</td>
<td>26.83</td>
</tr>
<tr>
<td>47, XXY</td>
<td>-1.609</td>
<td>-1.833</td>
<td>2720</td>
<td>16.11</td>
<td>19.58</td>
</tr>
<tr>
<td>45, XO</td>
<td>-0.545</td>
<td>-1.609</td>
<td>6100</td>
<td>10.22</td>
<td>15.31</td>
</tr>
<tr>
<td>Trisomy 9</td>
<td>-0.287</td>
<td>-0.635</td>
<td>2510</td>
<td>18.09</td>
<td>5.11</td>
</tr>
<tr>
<td>-18 + der 18</td>
<td>-0.233</td>
<td>-0.545</td>
<td>2460</td>
<td>12.58</td>
<td>4.52</td>
</tr>
<tr>
<td>46, XY 7q+</td>
<td>0.247</td>
<td>0.140</td>
<td>1480</td>
<td>0.47</td>
<td>2.72</td>
</tr>
<tr>
<td>mosaic 46,XY</td>
<td>0.512</td>
<td>0.425</td>
<td>1110</td>
<td>2.36</td>
<td>4.03</td>
</tr>
<tr>
<td>47, XY +mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 8.6.
Interpretation algorithm for assessment of atypicality.
Table 8.12.
Summary of effect of screening using the interpretation algorithm in figure 8.6.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. with Down's risk &gt; 1:300</th>
<th>No. with Down's risk &gt; 1:300 but atypical of normal at quantile 5%</th>
<th>2%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trisomy 13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>n = 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trisomy 18</td>
<td>0</td>
<td>12</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>n = 21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>n = 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unaffected</td>
<td>78</td>
<td>183</td>
<td>73</td>
<td>43</td>
</tr>
<tr>
<td>n = 2000</td>
<td>(3.9%)</td>
<td>(9.2%)</td>
<td>(3.7%)</td>
<td>(2.2%)</td>
</tr>
</tbody>
</table>
Figure 8.7.
Diagram to show graphically the basis of atypicality checking. The ellipses represent the 99% contours of normality and Down syndrome. Circles are trisomy 18 cases and squares are trisomy 13 cases.
also obvious that a greater proportion of unaffected results are flagged as atypical than the $\chi^2$ distribution cut offs should flag. This discrepancy can be explained by the failure of the MoM distribution to be completely Gaussian (chapter 4, section 4.3.3).

In conclusion, assessment of atypicality does not replace the likelihood ratio method for detection of Down syndrome but it may be used to aid detection of other anomalies by systematising identification of odd results. Adding atypicality testing to screening should not affect detection rates for Down syndrome provided the risk cut offs are not altered. The effect on false positive rates will depend on the action taken when an atypical result is given. There are several centres in the UK whose software allows assessment of atypicality. In those centres which use the facility, the general response to an atypical result is to indicate a reassessment of gestation age and if dating is correct to continue to a detailed 'anomaly' scan. Further investigation and management then depends on the judgement of the responsible obstetrician.

8.4.2 An examination of the discriminant function method of Nørgaard-Pedersen et al, (1990)

The triple-analyte discriminant function method for screening for Down syndrome described by Nørgaard-Pedersen has already been described in chapter 3, section 3.2.4.2. In addition to this a 'double' analyte discriminant function was described such that:

\[
D = \text{AFP} - 0.448 \times \text{HCG}
\]
In examining the effectiveness of the discriminant function method, the most important first step is to decide what the question is: As previously stated in section 8.2.3 it would be of dubious validity to define a new set of discriminant function parameters unless a different set of data was available to test them against. However, the question that is of most importance is whether the discriminant function method is transferable: i.e. whether the published parameters could be used in a laboratory that is setting up a Down syndrome screening program and therefore does not have its own set of Down syndrome parameters.

To answer this question, the same set of 52 Down syndrome cases and 536 unaffected results that were evaluated in section 8.2.3 were examined. 'Triple' and 'double' discriminant functions (equations 3.3 and 8.2) were calculated for all sets of results and were numerically integrated with a range of maternal age risks using the maternal age risk calculation of Tabor et al., (1987) (equation 3.4). These results were integrated with the age distribution of pregnancies in the UK to allow determination of false positive and detection rates for a series of cut off values. The rates so determined are shown in table 8.13. These are shown as a ROC plot in figure 8.8 in comparison with the ROC plot for the same data tested by the discriminant function method using a 'double' test and the Wald et al., (1988a) population parameters (the ROC plot for the 'triple' test using the 1988a parameters is not shown because this was not significantly different from that for the 'double' test).

From the graph it is obvious that the discriminant function test method gives inferior results when compared to the likelihood ratio method. This does not mean that the discriminant function test is useless
Table 8.13
Detection and false positive rates for 'double' and 'triple' testing using the discriminant function method of Nørgaard-Pedersen (1990).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>double test</td>
<td>double test</td>
</tr>
<tr>
<td></td>
<td>FPR</td>
<td>DR</td>
</tr>
<tr>
<td>1: 100</td>
<td>1.67</td>
<td>39.56</td>
</tr>
<tr>
<td></td>
<td>- 0.90</td>
<td>0.86</td>
</tr>
<tr>
<td>1: 150</td>
<td>2.70</td>
<td>45.88</td>
</tr>
<tr>
<td></td>
<td>- 0.70</td>
<td>1.49</td>
</tr>
<tr>
<td>1: 200</td>
<td>3.88</td>
<td>51.02</td>
</tr>
<tr>
<td></td>
<td>- 0.55</td>
<td>3.27</td>
</tr>
<tr>
<td>1: 250</td>
<td>4.88</td>
<td>54.90</td>
</tr>
<tr>
<td></td>
<td>- 0.45</td>
<td>6.08</td>
</tr>
<tr>
<td>1: 300</td>
<td>5.81</td>
<td>58.70</td>
</tr>
<tr>
<td></td>
<td>- 0.35</td>
<td>11.87</td>
</tr>
<tr>
<td></td>
<td>- 0.30</td>
<td>15.84</td>
</tr>
</tbody>
</table>
Figure 8.8
ROC plot of data in table 8.13. Circles represent ROC plot for likelihood ratio testing using Wald et al, 1988a. Squares represent 'triple' discriminant function testing and triangles represent 'double' discriminant function testing.
but rather implies that it is more data dependant than the likelihood ratio method, i.e. the effectiveness of the discriminant function is enhanced when the data that is being tested is closely related to the data the function was designed from. Thus although the same analytes are used to derive the risk, the failure of MoMs to act as a common currency means that the discriminant function is less efficient. Thus, this method of deriving Down syndrome risks is less suitable than the likelihood ratio method because it cannot be transferred from one laboratory to another without recalculation of the parameters: Since derivation of parameters needs a significant data set which is not likely to be available for the affected cases, the discriminant function method is not the method of first choice.

8.4.3 An examination of the method of Crossley et al, 1991

In 1991, Crossley et al described a method for two analyte screening where AFP and HCG MoMs were converted to a single variable (the HCG : AFP ratio) which was then used after log transformation in a univariate likelihood ratio format for calculation of Down syndrome risk. The population means and standard deviations described by Crossley et al were 0.0043 ± 0.2714 (normals) and 0.4502 ± 0.3046 (Down syndrome). To evaluate whether this method is effective (using the same premise as in section 8.4.2, i.e. whether it will work for a lab setting up a screening program) the same 52 cases of Down syndrome that were used in section 8.4.2 and 5082 unaffected pregnancy results were evaluated using the likelihood ratio method using population parameters from Wald et al, 1988a and using the method of Crossley et al.
In the Crossley paper the HCG : AFP ratios for the 410 control and 82 abnormal samples was tested for adherence to a LOG Gaussian distribution with a Kolgomorv-Smirnov test and shown to be acceptable. The distribution of the unaffected samples described above is shown in table 8.14. When tested for LOG Gaussian behavior by probability plotting and by the Liliefors test, however, both tests suggested a Non-Gaussian distribution (Liliefors test statistic 0.0205: cut offs: $\omega_{0.95} = 0.0124$, $\omega_{0.99} = 0.0145$). Table 8.15 shows the detection and false positive rates derived by numerical integration of likelihood ratios with the age distribution as described above. To illustrate this more clearly, figure 8.9 is a ROC plot comparing the performance of the likelihood ratio method using AFP and HCG as independent variables with the performance of the Crossley et al, method.

It is clear from the ROC plot that the HCG : AFP ratio method is less effective than the bivariate likelihood ratio method. This has also been described by other researchers (Cuckle et al, 1989: Spencer et al, 1992). Cuckle et al, responding to a suggestion by Arab et al, (1988) that a simple HCG : AFP ratio without consideration of maternal age was sufficient, showed that a decrease in detection of up to 17% resulted from the ratio method. Spencer et al, showed that for AFP + free $\beta$ HCG screening, use of the ratio method caused an approximately 5% decrease in detection. Figure 8.10 shows a comparison of the ratio method with the Nørgaard-Pedersen 'triple' discriminant function test method which shows that the ratio method is more effective.
Table 8.14  
Population Distribution for HCG / AFP ratio derived from 5082 results from the Gwent screening program.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>HCG / AFP</th>
<th>LOG(_{10}(\text{HCG/ AFP}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.205</td>
<td>-0.688</td>
</tr>
<tr>
<td>2.5</td>
<td>0.278</td>
<td>-0.556</td>
</tr>
<tr>
<td>5</td>
<td>0.348</td>
<td>-0.458</td>
</tr>
<tr>
<td>10</td>
<td>0.451</td>
<td>-0.346</td>
</tr>
<tr>
<td>15</td>
<td>0.537</td>
<td>-0.270</td>
</tr>
<tr>
<td>20</td>
<td>0.613</td>
<td>-0.213</td>
</tr>
<tr>
<td>25</td>
<td>0.679</td>
<td>-0.168</td>
</tr>
<tr>
<td>30</td>
<td>0.743</td>
<td>-0.129</td>
</tr>
<tr>
<td>35</td>
<td>0.807</td>
<td>-0.093</td>
</tr>
<tr>
<td>40</td>
<td>0.875</td>
<td>-0.058</td>
</tr>
<tr>
<td>45</td>
<td>0.938</td>
<td>-0.028</td>
</tr>
<tr>
<td>50</td>
<td>1.020</td>
<td>0.009</td>
</tr>
<tr>
<td>55</td>
<td>1.097</td>
<td>0.040</td>
</tr>
<tr>
<td>60</td>
<td>1.181</td>
<td>0.072</td>
</tr>
<tr>
<td>65</td>
<td>1.278</td>
<td>0.107</td>
</tr>
<tr>
<td>70</td>
<td>1.390</td>
<td>0.143</td>
</tr>
<tr>
<td>75</td>
<td>1.513</td>
<td>0.180</td>
</tr>
<tr>
<td>80</td>
<td>1.673</td>
<td>0.223</td>
</tr>
<tr>
<td>85</td>
<td>1.868</td>
<td>0.271</td>
</tr>
<tr>
<td>90</td>
<td>2.139</td>
<td>0.330</td>
</tr>
<tr>
<td>95</td>
<td>2.676</td>
<td>0.427</td>
</tr>
<tr>
<td>97.5</td>
<td>3.243</td>
<td>0.511</td>
</tr>
<tr>
<td>99</td>
<td>3.904</td>
<td>0.592</td>
</tr>
</tbody>
</table>
Table 8.15
Detection and false positive rates for screening using the method of Crossley et al.

<table>
<thead>
<tr>
<th>Risk Cut off</th>
<th>False Positive Rate (%)</th>
<th>Detection Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: 50</td>
<td>0.87</td>
<td>21.06</td>
</tr>
<tr>
<td>1: 100</td>
<td>1.67</td>
<td>30.99</td>
</tr>
<tr>
<td>1: 150</td>
<td>2.60</td>
<td>35.18</td>
</tr>
<tr>
<td>1: 200</td>
<td>3.52</td>
<td>37.95</td>
</tr>
<tr>
<td>1: 250</td>
<td>4.51</td>
<td>42.03</td>
</tr>
<tr>
<td>1: 300</td>
<td>5.67</td>
<td>46.09</td>
</tr>
<tr>
<td>1: 350</td>
<td>6.69</td>
<td>49.67</td>
</tr>
<tr>
<td>1: 400</td>
<td>7.91</td>
<td>53.96</td>
</tr>
<tr>
<td>1: 450</td>
<td>9.19</td>
<td>57.72</td>
</tr>
<tr>
<td>1: 500</td>
<td>10.09</td>
<td>60.16</td>
</tr>
</tbody>
</table>
Figure 8.9
Figure 8.10
Comparison of the ratio method of Crossley et al (circles) with the discriminant function method of Nørgaard-Pedersen et al (triangles).
8.4.4 An examination of the method of Muller et al, 1990

In 1990, Muller and Boué described a simple method for Down syndrome screening where all patients whose HCG result exceeded a defined cut off were offered amniocentesis. The cut offs were defined by the percentiles of the HCG distribution. By definition therefore, the false positive rate can be derived from the cut off: i.e. if the cut off is the 95%ile, the false positive rate is 100 - 95 % = 5%. Table 8.16 shows the detection rates for a selection of false positive rates for the Muller and Boué method evaluated with the 52 Down syndrome cases described above. Since these rates are based solely on the HCG level with no contribution for maternal age, it is not possible to integrate against the age distribution as in sections 8.4.2 and 8.4.3. The result of this is that the ROC plot of this data (figure 8.11) is not smoothed when compared to the earlier ROC plots. Once again, it is obvious that the likelihood ratio method is superior. Figure 8.12 shows the comparison of the HCG cut off method with the HCG : AFP ratio method. The lack of smoothing of the ROC plot for the HCG cut off method makes interpretation difficult but it can be concluded that the efficiency of the two methods is quite similar.

8.4.5 Conclusions

Any laboratory starting a new Down syndrome screening program will not have data on sufficient Down syndrome cases to derive its own population parameters for the Down syndrome group. It will therefore be necessary to use published parameters to set up their risk calculation algorithm. On the basis of the comparison above, it appears that the
Table 8.16
Detection and false positive rates for screening using the method of Muller and Boué.

<table>
<thead>
<tr>
<th>Percentile Cut off (%)</th>
<th>HCG MoM</th>
<th>False Positive Rate (%)</th>
<th>Detection Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>3.461</td>
<td>1</td>
<td>26.9</td>
</tr>
<tr>
<td>98</td>
<td>3.015</td>
<td>2</td>
<td>34.6</td>
</tr>
<tr>
<td>97</td>
<td>2.735</td>
<td>3</td>
<td>40.3</td>
</tr>
<tr>
<td>96</td>
<td>2.510</td>
<td>4</td>
<td>42.3</td>
</tr>
<tr>
<td>95</td>
<td>2.341</td>
<td>5</td>
<td>42.3</td>
</tr>
<tr>
<td>94</td>
<td>2.246</td>
<td>6</td>
<td>44.2</td>
</tr>
<tr>
<td>93</td>
<td>2.160</td>
<td>7</td>
<td>46.2</td>
</tr>
<tr>
<td>92</td>
<td>2.090</td>
<td>8</td>
<td>50.0</td>
</tr>
<tr>
<td>91</td>
<td>2.026</td>
<td>9</td>
<td>51.9</td>
</tr>
<tr>
<td>90</td>
<td>1.969</td>
<td>10</td>
<td>55.7</td>
</tr>
</tbody>
</table>
Figure 8.11
Figure 8.12
Comparison of the ratio method of Crossley et al (circles) with the HCG cut off method of Muller and Boué (triangles).
likelihood ratio method will offer the best detection rate for any given false positive rate. When sufficient local data has been collected to allow determination of population parameters optimal for the assays used by the laboratory it may be possible to use the other methods of determining who should be offered amniocentesis effectively.

8.5 The Effect of Weight Correction on Detection (Reynolds, 1993)

To determine whether weight correction improves detection in Down syndrome screening, 24 cases of Down syndrome pregnancy for which the maternal weight was known were examined. Initially, Down syndrome risks were calculated using uncorrected MoMs and with weight corrected MoMs. These results are shown in table 8.17. After weight correction 3 extra cases were detected (i.e. 13 cases (54%) detected before correction, 16 cases (66.7%) after correction). This cannot be evaluated without considering the effect of weight correction on false positive rates. This was evaluated by examining 2857 unaffected pregnancies for a variety of risk cut offs (table 8.18). This shows that there is a reduction in the false positive rate when weight correction is applied. To conclusively demonstrate whether weight correction is of benefit, a ROC plot is useful: this was generated by taking the likelihood ratios for uncorrected and weight corrected results and numerically integrating them against a grid of maternal ages as previously described (section 8.2.3). The ROC plot so derived is shown as figure 8.13 and demonstrates an approximately +5% increase in detection for a constant false positive rate. This is an extremely small study and as such in no way provides conclusive evidence that weight correction is useful but it is suggestive that it may provide a significant improvement in detection.
Table 8.17
Comparison of uncorrected and weight corrected MoMs in Down syndrome detection.

<table>
<thead>
<tr>
<th>Maternal Age</th>
<th>Maternal Weight (kg)</th>
<th>AFP MoM</th>
<th>HCG MoM</th>
<th>Uncorr’ed Risk</th>
<th>Corrected Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>39.1</td>
<td>84.6</td>
<td>0.5</td>
<td>5.5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>45.8</td>
<td>71.7</td>
<td>1.6</td>
<td>3.5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>40.0</td>
<td>62.5</td>
<td>0.9</td>
<td>3.1</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>35.3</td>
<td>56.3</td>
<td>0.6</td>
<td>3.7</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>35.7</td>
<td>82.0</td>
<td>0.5</td>
<td>3.2</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>38.7</td>
<td>86.6</td>
<td>0.8</td>
<td>2.7</td>
<td>33</td>
<td>31</td>
</tr>
<tr>
<td>40.0</td>
<td>67.5</td>
<td>0.7</td>
<td>2.1</td>
<td>36</td>
<td>33</td>
</tr>
<tr>
<td>34.9</td>
<td>77.5</td>
<td>1.1</td>
<td>3.9</td>
<td>47</td>
<td>44</td>
</tr>
<tr>
<td>42.8</td>
<td>62.4</td>
<td>0.5</td>
<td>0.7</td>
<td>90</td>
<td>76</td>
</tr>
<tr>
<td>34.4</td>
<td>57.5</td>
<td>1.0</td>
<td>3.0</td>
<td>96</td>
<td>90</td>
</tr>
<tr>
<td>36.3</td>
<td>67.2</td>
<td>0.8</td>
<td>1.7</td>
<td>191</td>
<td>177</td>
</tr>
<tr>
<td>36.1</td>
<td>71.8</td>
<td>0.8</td>
<td>1.7</td>
<td>189</td>
<td>189</td>
</tr>
<tr>
<td>29.5</td>
<td>72.6</td>
<td>1.0</td>
<td>2.8</td>
<td>259</td>
<td>246</td>
</tr>
<tr>
<td>39.4</td>
<td>45.5</td>
<td>0.8</td>
<td>0.9</td>
<td>357</td>
<td>254</td>
</tr>
<tr>
<td>31.6</td>
<td>61.0</td>
<td>0.4</td>
<td>1.2</td>
<td>303</td>
<td>259</td>
</tr>
<tr>
<td>23.4</td>
<td>59.5</td>
<td>0.9</td>
<td>2.8</td>
<td>318</td>
<td>295</td>
</tr>
<tr>
<td>30.4</td>
<td>80.1</td>
<td>0.2</td>
<td>0.5</td>
<td>373</td>
<td>411</td>
</tr>
<tr>
<td>23.9</td>
<td>54.0</td>
<td>1.0</td>
<td>2.3</td>
<td>636</td>
<td>573</td>
</tr>
<tr>
<td>33.1</td>
<td>66.1</td>
<td>1.1</td>
<td>1.7</td>
<td>646</td>
<td>547</td>
</tr>
<tr>
<td>29.2</td>
<td>51.8</td>
<td>0.7</td>
<td>1.3</td>
<td>974</td>
<td>785</td>
</tr>
<tr>
<td>26.6</td>
<td>94.5</td>
<td>1.2</td>
<td>2.0</td>
<td>1086</td>
<td>1070</td>
</tr>
<tr>
<td>28.4</td>
<td>63.2</td>
<td>0.6</td>
<td>0.9</td>
<td>1695</td>
<td>1466</td>
</tr>
<tr>
<td>18.9</td>
<td>68.4</td>
<td>0.6</td>
<td>1.0</td>
<td>1972</td>
<td>1811</td>
</tr>
<tr>
<td>20.7</td>
<td>73.2</td>
<td>1.1</td>
<td>1.2</td>
<td>3989</td>
<td>3818</td>
</tr>
</tbody>
</table>
Table 8.18
The effect of weight correction on false positive rates

<table>
<thead>
<tr>
<th>Risk cut off</th>
<th>False Positive Rate (%)</th>
<th>Not Weight corrected</th>
<th>Weight Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: 50</td>
<td></td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>1: 100</td>
<td></td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>1: 150</td>
<td></td>
<td>1.86</td>
<td>1.64</td>
</tr>
<tr>
<td>1: 200</td>
<td></td>
<td>3.04</td>
<td>2.63</td>
</tr>
<tr>
<td>1: 250</td>
<td></td>
<td>4.48</td>
<td>3.85</td>
</tr>
<tr>
<td>1: 300</td>
<td></td>
<td>5.46</td>
<td>5.08</td>
</tr>
<tr>
<td>1: 350</td>
<td></td>
<td>6.93</td>
<td>6.19</td>
</tr>
<tr>
<td>1: 400</td>
<td></td>
<td>8.01</td>
<td>7.45</td>
</tr>
<tr>
<td>1: 450</td>
<td></td>
<td>8.96</td>
<td>8.12</td>
</tr>
<tr>
<td>1: 500</td>
<td></td>
<td>10.36</td>
<td>9.21</td>
</tr>
</tbody>
</table>
Figure 8.13
Comparison of ROC plots for uncorrected (circles) and weight corrected (triangles) data.
8.6 Conclusions

1) The likelihood ratio method of risk determination includes calculation of the maternal age related risk. The result of this is that the efficiency of screening is affected by maternal age with a lower detection and false positive rate in younger women. This factor may be of great importance when counselling before testing or after giving a 'high risk' result.

2) A wide variety of different population parameters for Down syndrome and unaffected groups have been reported in the literature. Comparison of two published data sets with locally derived data showed that in computer simulations there were significant differences in the predicted screening performances. When the same data sets were tested with 'real' data, the performance was totally different to that predicted. The lesson of this is that computer simulations must be evaluated critically before the output can be regarded as 'truth'. The only real test of whether a screening protocol is effective is to test against real patient data.

3) There is a great deal of debate about which analytes should be used in Down screening. The data presented here suggests that adding unconjugated oestriol to a 'double' test does not add detection but other centres, notably Cambridge, find that it adds significantly (M. Ferguson-Smith, answers to questions at conference on Down syndrome, London, May 1993). It is difficult to explain why this should be but it may be that the same factor that caused loss of detection when using the Wald et al., 1992a population parameters for uE₃ in section 8.2.3 above apply in Cambridge to a different analyte.
According to statistical theory, there should always be a benefit from adding an extra analyte and analytical imprecision does not affect this (D. Wright, Personal communication: H. Cuckle, answers to questions at conference on Down syndrome, London, May 1993). The alternative point of view is that analytical imprecision is important and therefore the number of contributing analytes is limited by the precision of the assays. These different viewpoints are difficult to resolve but are both correct! From a statisticians’ viewpoint, the screening process is looking at the whole population and anything that increases the discrimination between the sub-groups is useful meaning that analytical imprecision is not important. From a clinicians’ viewpoint, the population is immaterial and it is the individual that is important and therefore the clinician wants to give as accurate a risk as possible to each woman and therefore analytical imprecision is critical. These views are impossible to reconcile and therefore to determine which is the best combination of markers and whether a 'double' or a 'triple' test is required, further research involving large prospective trials will be needed.

4) The data presented here suggests that for a laboratory starting a new Down syndrome screening program, the best method of calculating Down syndrome risks is the likelihood ratio method.

5) Other maternal factors may be important: correction for maternal weight offers a small improvement in detection.
chapter 9:

EVALUATION OF DOWN SYNDROME SCREENING IN GWENT AND SOUTH GLAMORGAN

9.1 PREDICTION OF DOWN SYNDROME INCIDENCE

9.2 ONE YEAR'S EXPERIENCE OF DOWN SYNDROME SCREENING

9.2.1 Description of Study
9.2.2 Results of Study
9.2.3 Discussion
9.2.4 Inter-Disciplinary Concerns About Screening
9.1 Prediction of Down Syndrome Incidence

Before the effectiveness of a screening program can be evaluated it is important to have some idea of the expected performance. Since the age distribution of the pregnant population of Gwent is known (table 4.1) and the risk of a Down syndrome pregnancy related to age is known (section 3.2.1), it is possible to estimate the expected number of Down syndrome cases per year. The combined birth rate for Gwent and the University Hospital of Wales is approximately 10,000 births per year. Table 9.1 shows the derivation of the expected number of Down syndrome births based on this total. Assuming no prenatal detection, a total of 13.3 Down syndrome affected individuals would be expected to be born each year.

However, the expected number of Down syndrome affected births is an under estimate of the number of Down syndrome affected fetuses that are 'available for detection' in the second trimester. It is estimated that there is a spontaneous abortion rate of 29% for Down syndrome fetuses between the second trimester and term (Hook et al, 1983; Palomaki and Haddow, 1987). Therefore, a corrected estimate of the number of Down syndrome affected pregnancies at the time of screening for the UHW / Gwent population is 18.76.
Table 9.1. Prediction of expected number of Down syndrome births per year for Gwent and the University Hospital of Wales assuming no prenatal detection. Down births = number of Down syndrome cases if total number of births = 10000.

<table>
<thead>
<tr>
<th>Age years</th>
<th>Frequency (%)</th>
<th>p(Downs) (%)</th>
<th>Down births</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>0.0186</td>
<td>0.0631</td>
<td>0.00117</td>
</tr>
<tr>
<td>14</td>
<td>0.0000</td>
<td>0.0632</td>
<td>0.00</td>
</tr>
<tr>
<td>15</td>
<td>0.0186</td>
<td>0.0634</td>
<td>0.00179</td>
</tr>
<tr>
<td>16</td>
<td>0.5389</td>
<td>0.0636</td>
<td>0.03427</td>
</tr>
<tr>
<td>17</td>
<td>0.7619</td>
<td>0.0639</td>
<td>0.04869</td>
</tr>
<tr>
<td>18</td>
<td>1.6911</td>
<td>0.0642</td>
<td>0.10857</td>
</tr>
<tr>
<td>19</td>
<td>2.6947</td>
<td>0.0648</td>
<td>0.17462</td>
</tr>
<tr>
<td>20</td>
<td>3.7725</td>
<td>0.0654</td>
<td>0.24672</td>
</tr>
<tr>
<td>21</td>
<td>3.7911</td>
<td>0.0663</td>
<td>0.25135</td>
</tr>
<tr>
<td>22</td>
<td>5.4451</td>
<td>0.0675</td>
<td>0.36754</td>
</tr>
<tr>
<td>23</td>
<td>6.1884</td>
<td>0.0691</td>
<td>0.42762</td>
</tr>
<tr>
<td>24</td>
<td>6.4486</td>
<td>0.0712</td>
<td>0.45914</td>
</tr>
<tr>
<td>25</td>
<td>7.7123</td>
<td>0.0740</td>
<td>0.57071</td>
</tr>
<tr>
<td>26</td>
<td>8.4928</td>
<td>0.0777</td>
<td>0.66000</td>
</tr>
<tr>
<td>27</td>
<td>8.4557</td>
<td>0.0828</td>
<td>0.70013</td>
</tr>
<tr>
<td>28</td>
<td>7.6937</td>
<td>0.0893</td>
<td>0.68705</td>
</tr>
<tr>
<td>29</td>
<td>6.9875</td>
<td>0.0982</td>
<td>0.68617</td>
</tr>
<tr>
<td>30</td>
<td>6.6902</td>
<td>0.1098</td>
<td>0.73458</td>
</tr>
<tr>
<td>31</td>
<td>4.8132</td>
<td>0.1256</td>
<td>0.60454</td>
</tr>
<tr>
<td>32</td>
<td>4.1256</td>
<td>0.1464</td>
<td>0.60399</td>
</tr>
<tr>
<td>33</td>
<td>3.0478</td>
<td>0.1742</td>
<td>0.53093</td>
</tr>
<tr>
<td>34</td>
<td>2.6017</td>
<td>0.2105</td>
<td>0.54766</td>
</tr>
<tr>
<td>35</td>
<td>2.0256</td>
<td>0.2604</td>
<td>0.52747</td>
</tr>
<tr>
<td>36</td>
<td>1.8398</td>
<td>0.3246</td>
<td>0.59720</td>
</tr>
<tr>
<td>37</td>
<td>1.3195</td>
<td>0.4135</td>
<td>0.54561</td>
</tr>
<tr>
<td>38</td>
<td>1.0593</td>
<td>0.5263</td>
<td>0.55751</td>
</tr>
<tr>
<td>39</td>
<td>0.7248</td>
<td>0.6860</td>
<td>0.49721</td>
</tr>
<tr>
<td>40</td>
<td>0.3717</td>
<td>0.8771</td>
<td>0.32601</td>
</tr>
<tr>
<td>41</td>
<td>0.2788</td>
<td>1.1723</td>
<td>0.32684</td>
</tr>
<tr>
<td>42</td>
<td>0.0743</td>
<td>1.4925</td>
<td>0.11089</td>
</tr>
<tr>
<td>43</td>
<td>0.1487</td>
<td>2.0456</td>
<td>0.30418</td>
</tr>
<tr>
<td>44</td>
<td>0.0929</td>
<td>2.5641</td>
<td>0.23820</td>
</tr>
<tr>
<td>45</td>
<td>0.0186</td>
<td>3.6306</td>
<td>0.06753</td>
</tr>
<tr>
<td>46</td>
<td>0.0186</td>
<td>4.8686</td>
<td>0.09056</td>
</tr>
<tr>
<td>47</td>
<td>0.0372</td>
<td>6.5631</td>
<td>0.24415</td>
</tr>
</tbody>
</table>

Total number of births predicted 13.31881
9.2 One Year’s Experience of Down Syndrome Screening
(Dawson et al., 1992, 1993)

9.2.1 Description of Study

When screening for Down syndrome was implemented in South Wales it was decided that it should initially be introduced as a pilot study to determine whether the screening method functioned as well as the predictions in the original description (Wald et al., 1988a) and whether it was a useful addition to routine antenatal care. The pilot study was set up as a collaborative venture between the biochemistry laboratories at the Royal Gwent Hospital (which also serves Nevill Hall Hospital, Abergavenny for the purposes of screening) and at the University Hospital of Wales; the Institute of Medical Genetics for Wales (which performs the cytogenetic analyses for all of South Wales); the obstetric departments of all three hospitals; and the Gwent Public Health Department.

At the UHW, a research midwife (Gillian Jones) was employed to carry out a cohort study of all women booking for antenatal care at the UHW during the period 23 February 1990 to 22 February 1991. During this period all women booking were assigned a trial number and positive ascertainment of all outcomes was made by examining case notes, Medical Genetics reports, delivery suite records and paediatric notes. The last women in the UHW cohort delivered in September 1991.

In Gwent, monthly listings of all ‘high risk’ results were printed by the laboratory and the obstetric outcomes for all of these cases were collected by Dr. M. Matharu of the Public Health Medicine Department as an extra arm of the obstetric audits already carried out (perinatal
mortality and outcome of pregnancies with high AFP). Any 'affected' outcomes of 'low risk' pregnancies were obtained from the reports of birth anomalies from obstetric and paediatric staff, and from cytogenetic reports from the Institute of Medical Genetics. This also allowed details of all amniocenteses to be collected. Data collection was carried out continuously from the implementation of the Down syndrome screening program in late 1989. Only that data relating to the same period as that covered by the UHW data is considered here. In addition to the performance of the screening program in terms of detection and false positive rates, the demands of screening on laboratory and clinical services was examined.

The two laboratories derived their own reporting protocols that suited the obstetric practices in the hospitals they covered. In Gwent, it was established that ultrasound checking of gestation dates was usually performed on the same day as blood sampling: requests for Down screening were made using the special form already described (figure 4.1) and results were supplied as a single risk estimate for the gestation date supplied to the laboratory. In the event of dating not being performed on the day of sampling and the estimated date being proved incorrect at a later appointment, the clinicians contacted the laboratory for a re-assessment of Down syndrome risk. At the UHW, dating was usually performed by crown-rump length estimation earlier in pregnancy but due to the established practice in NTD screening, risks were provided for weeks 15, 16, 17 and 18 with the risk corresponding to the supplied date highlighted. At all centres, a risk cut off of 1 in 300 was used to define which women should be offered amniocentesis. Both laboratories used essentially the same assays (as described in chapter 3).
The members of the study team were: Mr. A. Dawson and Gillian Jones (obstetrics dept., UHW), Dr. M. Matharu (Public Health dept., Gwent Health Authority), Dr. T. Reynolds and Rhys John (Biochemistry dept., UHW), Dr. M. Penney (Biochemistry dept., Royal Gwent Hospital), Dr. M. Creasy and Mr. P. Gregory (Institute of Medical Genetics, UHW).

9.2.1 Results of Study

The results of the study are shown in a tree diagram (fig. 9.1). In summary, the booking population was 9937 women, very close to the 10,000 expected. Based on this value, the estimate of expected Down syndrome cases at the time of screening can be corrected to 18.64 cases. Overall 93% of women were eligible for screening. The 7% who were ineligible were excluded because of late booking (gestation dates > 18 weeks), miscarriage prior to screening and multiple pregnancy. Of the 9283 eligible women, 91% accepted the offer of serum testing and received their result. There was a difference in the acceptance rates between the UHW and the Gwent hospitals with only 83% accepting screening at the UHW and 94% accepting in Gwent. The reasons for this are not clear but may have been due to differences between the populations of Gwent and South Glamorgan or to differences in counselling used by medical staff in the different centres.

Of the 8414 women who were screened, 3.5% had a calculated risk of greater than 1 in 300 with nearly identical proportions in Gwent and South Glamorgan. This similarity was expected since the assays were effectively identical, the medians and risk calculation method were identical and the populations of the two counties were very similar. In
Figure 9.1
Summary of prospective study of Down syndrome screening.
this 'high risk' group, the uptake of amniocentesis was 86% and 7 cases of Down syndrome were detected. Six of these were terminated but one woman opted to continue to term having arranged Social Services back up and forewarned all of her friends and relatives. Eight thousand one hundred and eighteen pregnancies were designated 'low risk' and amongst these women there were also 7 cases of Down syndrome. However, one of these 7 cases had a Down syndrome risk estimate of 1 in 303 and was offered amniocentesis with a subsequent termination. This could have been considered to be a detection by the screening program but due to the strict criteria set up for the study it has been counted as a false negative result albeit with a 'true positive' outcome. Amongst the women who were not serum tested there were 3 cases of Down syndrome, one of which was detected by amniocentesis carried out on the basis of maternal age. A list of all cases of Down syndrome during the study year is presented in table 9.2.

There were 17 cases of Down syndrome during the study period: This agrees quite well with the prediction of 18.64 in the second trimester and 13.3 at term. The detection rate in the screened population was 50% although if the one case that was detected with a risk of 1:303 is included this becomes 57%. It is interesting to note that the oldest woman in the screened group with a risk less than 1 in 300 was 34 years old and that all women aged greater than this who carried a Down syndrome fetus were identified. This is completely in keeping with the age-related properties of the screening test as described in chapter 8, section 8.1. It may be argued that those cases of Down syndrome in women in the 'high risk' group aged over 35 would have been detected without the serum screen because amniocentesis would have been carried out based on maternal
Table 9.2
Down syndrome cases identified during the study period. In the outcome column, TOP indicates termination of pregnancy, L/B indicates a live birth and S/B indicates a still birth.

<table>
<thead>
<tr>
<th>Site</th>
<th>Age @ EDD</th>
<th>Risk estimate</th>
<th>Outcome</th>
<th>Prenatal detection?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Screened: Risk &gt; 1:300</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHW</td>
<td>19</td>
<td>1 : 20</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td>UHW</td>
<td>40</td>
<td>1 : 36</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td>UHW</td>
<td>39</td>
<td>1 : 19</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td>Gwent</td>
<td>19</td>
<td>1 : 190</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td>Gwent</td>
<td>37</td>
<td>1 : 55</td>
<td>L/B</td>
<td>Yes</td>
</tr>
<tr>
<td>Gwent</td>
<td>37</td>
<td>1 : 286</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>1 : 99</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Screened: Risk &lt; 1:300</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHW</td>
<td>31</td>
<td>1 : 303</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td>UHW</td>
<td>32</td>
<td>1 : 1426</td>
<td>S/B</td>
<td>No</td>
</tr>
<tr>
<td>UHW</td>
<td>29</td>
<td>1 : 980</td>
<td>L/B</td>
<td>No</td>
</tr>
<tr>
<td>UHW</td>
<td>34</td>
<td>1 : 583</td>
<td>L/B</td>
<td>No</td>
</tr>
<tr>
<td>Gwent</td>
<td>22</td>
<td>1 : 1487</td>
<td>L/B</td>
<td>No</td>
</tr>
<tr>
<td>Gwent</td>
<td>29</td>
<td>1 : 773</td>
<td>L/B</td>
<td>No</td>
</tr>
<tr>
<td><strong>Not Screened</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHW</td>
<td>36</td>
<td>Amnio for age</td>
<td>TOP</td>
<td>Yes</td>
</tr>
<tr>
<td>UHW</td>
<td>40</td>
<td>Amnio for age</td>
<td>L/B</td>
<td>Yes</td>
</tr>
<tr>
<td>UHW</td>
<td>24</td>
<td>Late booker</td>
<td>L/B</td>
<td>No</td>
</tr>
</tbody>
</table>
age alone. However, studies at the UHW show that the uptake of amniocentesis when offered on the basis of age alone was: 16% for women aged 35 - 36 years; 35% for women aged 37-39 years and 46% for women age >= 40 years. This compares to the 86% uptake in women with a 'high risk' serum screen result. Furthermore, the historical detection rate for Down syndrome when amniocentesis was performed on the basis of maternal age alone was 1%.

During the study period 493 amniocenteses were performed on the booking population of 9937 women (4.96%). Of these, 253 were on women who had a 'high risk' result (detecting 1 case of Down syndrome per 36 amniocenteses) from their serum screen and 204 were on women who either had a 'low risk' result or who were ineligible for testing (detecting 1 case of Down syndrome per 80 amniocenteses). Overall there was 1 case of Down syndrome per amniocentesis. A breakdown of the amniocenteses during the study period is shown in table 9.3. Due to the data collection system in Gwent no age breakdown for population subgroups was available. The difference between the age distributions in the total populations of Gwent and UHW were compared using the $\chi^2$ test with a result of $\chi^2 = 0.678$ (P = Not Significant) implying that the amniocentesis rates are similar between the two centres. By examination of the sub-population figures, the proportions of amniocenteses in the 'high' and 'low risk' groups for both centres are very similar and it may be expected that the age breakdowns for the Gwent population would be similar to the UHW population.

During the study period, there were 150 amniocenteses carried out at the UHW of which approximately 50% were performed following a
Table 9.3
Amniocenteses performed during the study period

<table>
<thead>
<tr>
<th></th>
<th>Age &lt; 35</th>
<th>Age &gt;=35</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(screened and unscreened)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHW</td>
<td>83 (55%)</td>
<td>67 (45%)</td>
<td>150</td>
</tr>
<tr>
<td>Gwent</td>
<td>176 (51%)</td>
<td>167 (49%)</td>
<td>343</td>
</tr>
<tr>
<td>Total</td>
<td>259 (52%)</td>
<td>234 (48%)</td>
<td>493</td>
</tr>
</tbody>
</table>

|                  |          |          |       |
| **Screened population** |          |          |       |
| 'High' risk group   |          |          |       |
| UHW                | 49 (67%) | 24 (33%) | 73    |
| Gwent              | N/A      | N/A      | 180   |
| Total              | N/A      | N/A      | 253   |

|                  |          |          |       |
| **Screened population** |          |          |       |
| 'Low' risk group + others |          |          |       |
| UHW                | 34 (44%) | 43 (56%) | 77    |
| Gwent              | N/A      | N/A      | 163   |
| Total              | N/A      | N/A      | 240   |
'high' risk Down syndrome screening test result. This compared favourably with the previous year’s 154 amniocenteses before specific screening was available. The consistent level of amniocenteses was maintained by stopping performing them for investigation of elevated maternal serum AFP levels which were investigated by detailed ultrasound scanning, and a reduction in the number of women requesting amniocentesis for age alone.

9.2.3 Discussion

The screening test detected Down syndrome in 7 of the 14 cases that were detectable and could be considered to have lead to the detection of an 8th case. Of these 8 cases, 4 were in women aged less than 35 years which would not have been detected if screening was unavailable and the policy of amniocentesis for advanced maternal age had been continued. It may therefore be concluded that considerable benefits had resulted from the screening test. However, before the introduction of the test, concerns about Down syndrome had never been raised with those women aged less than 35 and there was therefore a potential for a new source of psychological stress for pregnant women in the study. These effects would be expected to be particularly significant in the 289 women given false positive results and the 6 women who were given negative results who delivered affected babies.

It would be ideal if the accuracy of prediction could be improved but it is unlikely that serum screening could provide the perfect screening test with 100% sensitivity and specificity. Overall the test does have advantages, the frequency of Down syndrome in the screened population
during the study period was 1 in 600; the risk in those who had a 'high' risk screening result was 1 in 42; and in those with a 'low' risk screening result was 1 in 1160, a 28-fold difference. In those women who were not screened, the overall risk was 1 in 508. Some of the difference between this group and the 1 in 600 rate in the screened women is due to the cases detected that would have miscarried naturally and would not have been otherwise detected.

The role of the screening test for women who do not wish to have a termination of pregnancy must not be underestimated: one woman in Gwent was able to receive counselling and paediatricians, social services, friends and relatives were able to be alerted and prepared for the impending birth of a Down syndrome affected child. The implications of screening therefore go beyond the cold economics of termination. A fuller evaluation of the health economic factors are beyond the scope of this study.

9.2.4 Inter-disciplinary concerns about screening

Following discussion amongst the members of the study team a number of conclusions and recommendations that were felt to be relevant to all Down syndrome screening programs were developed:

1) There appears to be a health gain from serum screening for Down syndrome in detecting a proportion of affected fetuses in a group of pregnancies where previously none were detected.

2) Gains accrue also in the women aged over 35 who avoid
amniocentesis.

3) In pregnancies where affected fetuses are terminated there is a nominal cost saving to the community services for assistance that would have been required following a Down’s birth (Gill et al, 1987).

4) The health gains are counter-balanced by concern about the effect of false positive and false negative results. Since gains and losses are difficult to assess in similar units the net gain is difficult to determine.

5) There are significant laboratory resource implications due to the introduction of Down syndrome screening: biochemistry laboratories must introduce new assays which may also result in an increase in MLSO staff requirements; cytogenetic resource requirements may be increased or may remain constant due to a decrease in amniocenteses for NTD screening and for increased maternal age (the precise implications depend on local conditions but may include a significant increase in amniocentesis in districts where ultrasound is the main assessment method used when there is a significantly elevated maternal serum AFP concentration).

6) There is a significant increase in obstetric and midwifery service workload due to the need to counsel women before testing and after a positive result. The effects on amniocentesis workload will depend on local conditions as for the cytogenetic service.

7) In addition to one-to-one counselling, clear written information
should be developed to suit local needs. Where this requires translation into minority languages this must be considered.

8) Screening performance must be audited as an on-going process to ensure that the program is performing adequately. This means that there is a minimum size for a screening program which should be enough for a significant number of cases of Down syndrome to occur each year. A suggested minimum population size would be 5000 births per year ensuring approximately 10 cases of Down syndrome per year.

9) For the psychological well-being of the screening population and to ensure efficient audit, all necessary services (including biochemistry, counselling, amniocentesis and cytogenetics) should be available locally and should be geared up to provide Down syndrome screening. Therefore there should be good links between all of the units concerned. For this reason, use of remote laboratories with no direct communications with local services should be avoided as far as possible.

10) Since there were a significant number of women who were not screened because they booked too late, education about screening is required to emphasis the need for early diagnosis of pregnancy and early booking.
CONCLUSIONS AND PROSPECTS FOR THE FUTURE

10.1 DISCUSSION

10.1.1 Why should we screen for Down syndrome?
10.1.2 What method of screening should we use?
10.1.3 Which analytes should we use?
10.1.4 How should we standardise for gestation age?
10.1.5 What other maternal factors should be taken into account?
10.1.6 What is the significance of gestation dating?
10.1.7 Which mathematical algorithm should we use?
10.1.8 What can we expect from Down syndrome screening?
10.1.9 What are the concerns about Down syndrome screening?

10.2 FUTURE AVENUES IN SERUM SCREENING

10.2.1 Correlation between analytes may not be bad.
10.2.2 Correlation between analytes may be spurious.
10.2.3 Gestation dating may affect medians.
10.2.4 Risk calculation algorithms.
10.2.5 Ethnic differences.
10.2.6 Inter-kit variations in population parameters.
10.2.7 Within-kit variability
10.2.8 Within patient variation

10.3 FUTURE AVENUES IN DOWN SYNDROME SCREENING

10.3.1 Fetal cells in the maternal circulation.
10.3.2 FISHing for a diagnosis.
10.3.3 PCR possibilities.
10.1 Discussion

10.1.1 Why should we screen for Down syndrome?

In the absence of screening, approximately 1 in 700 babies will be born with Down syndrome, the exact rate being partly determined by the age distribution of the pregnant population. Down syndrome forms the largest single group of chromosomal abnormalities and as a result is also said to be the single largest cause of mental handicap. Furthermore those Down syndrome individuals who do not suffer from congenital heart disease have a life expectancy similar to that of normal individuals although they also have a higher rate of premature dementia.

These factors mean that from an economic standpoint it is viable to prevent the birth of Down syndrome babies because of the considerable saving that can be made by avoiding the need to care for a mentally handicapped child and also preventing the decrease in parental productivity that such a child would cause. However, if economics were the only factor, screening would not be ethically acceptable. The other factor that has a major effect is 'consumer demand': i.e. screening for Down syndrome is acceptable to the general public. Most pregnant women when asked whether they want a boy or a girl will reply that they don't mind so long as it's normal. This is reinforced because many of them have seen the problems that a handicapped child can cause. It was this factor that was behind the acceptance of NTD screening when it was introduced in the 70's despite the fact that many NTD affected babies are of normal intelligence. In the case of Down syndrome the outlook is for a child that will be dependent until its parents die, an event that may be presaged by the worry of what is to happen to the child when it is left
The extent of 'consumer demand' can be gauged by the Cardiff experience where initially due to financial constraints, screening was only offered to women over 25 years of age but within 6 months this had to be extended to all women. Younger women discussing their pregnancies with other women found that the test was available and demanded that it be provided for them. The extent of the demand can be seen in figure 9.2 where the results of our prospective trial are shown: over 90% of those women eligible for the test accepted it and in those receiving a 'high' risk result, 85% accepted amniocentesis.

The answer to the question, why screen for Down syndrome can therefore be broken into two parts: Firstly, there is overwhelming pressure from the public to provide the service now that the technology is available; Secondly, for economic reasons.

10.1.2 What method of screening should we use?

In deciding the method for Down syndrome screening, there are at least 2 questions that need to be answered but they cannot be answered in isolation. The most important questions are: what is the best method? and what can we afford? The answer to the first question is obviously the method with the highest possible detection rate but this must be tempered by the second question. In a cash-unlimited situation, it would be possible to achieve 100% detection by giving amniocentesis to all women. Apart from the excessive cost, however this would be rejected because of the rate of losses of normal fetuses. Since we are in a cash-
limited environment however, we must achieve the maximum detection at the lowest possible false positive rate. This can only be done by using the serum screen as a preliminary test to determine who should be offered amniocentesis.

10.1.3 Which analytes should we use?

The decision on which analytes should be used depends on when screening is to be performed. If this is in the first trimester the current combination of choice is AFP with free $\beta$ HCG but this is still subject to confirmation and may be improved upon by new discoveries. PAPP-A has been suggested to be an improvement over AFP. In the second trimester, other screening modalities must also be considered. Due to its utility for NTD screening, the first marker used by most laboratories is AFP. Second is HCG (either total or free $\beta$) although free $\beta$ is rapidly appearing to be superior (Spencer et al, 1993b). The third marker is uE$_3$ and there may be other markers that could be added in a 4th place. The rationale being that AFP is used for its detection of NTD’s and this means that the second marker must be HCG (as the combination of AFP with uE$_3$ is not very efficient). Whether a third marker is used depends on funding and whether the lab believes it adds anything.

The decision about whether to use uE$_3$ or not depends partly on the viewpoint of the person making the decision: the statistician’s view is that adding extra analytes must provide extra benefit because extra information will always increase discrimination between different groups in the population; the clinician’s view is that whilst extra analytes may increase discrimination between populations, the focus of the test is an
individual and therefore the result for that one person should be as correct an estimate as possible. Thus if adding extra analytes reduces the confidence that can be placed in the result, it may be counter-productive to add the extra analyte.

The number of analytes may however be increased if the precision consequences are dealt with: several multiple-analyte assays are now in development (Macri et al, 1992 (ELISA method): Pettersson et al, 1993 (DELFIA method (Wallac Ltd.,))) which means that overall precision of risk estimates may be improved. Once precision has been dealt with we return to the second factor that determines whether extra analytes can be used: cost. It must be proven that the extra detection arising from adding an extra analyte does not increase costs beyond an acceptable limit or if the global money pool is constant, an increase in serum analytes means that less amniocenteses can be performed which means that the detection at the new lesser acceptable false positive rate must be considered.

10.1.4 How should we standardise for gestational age?

The need to standardise for gestation age is well-known and at present the only method available is to use multiples of the median. As shown in section 7.1.1, the MoM is far from being the common currency that it is assumed to be but a replacement is hard to come by. Until such a time arrives that a better method is available we must continue to use the MoM and must therefore ensure that it is calculated in the most effective way possible. 

Not all methods of deriving the medians for calculation of MoMs
are equally valid. The ideal population distribution for a MoM converted variable should be centred on 1.0 and should be transformable to a Gaussian pattern (currently achieved by LOG conversion). The only methods that satisfy these requirements are weighted regression of LOG transformed weekly derived medians (after Wald et al, 1988a) and regression of date against log transformed results. Even after LOG transformation the population distribution of MoMs does not completely satisfy the Gaussian model especially in the tails. We should therefore continue to strive towards a replacement for the MoM.

10.1.5 What other maternal factors should be taken into account?

For many years in NTD screening, it has been the practice to correct the AFP MoM for maternal weight as this was shown to improve detection rates without increasing false positive rates. In Down syndrome screening, the effect of maternal weight would be expected to have a lesser contribution because if correction of AFP results in a lower MoM (indicating a higher risk result), the MoM for HCG will also be corrected downwards (indicating a lower risk). Thus the two effects partially cancel each other out. Despite this, there is a small increase in detection when weight correction is performed (Fig 8.13).

A variable related to weight is height. This was also shown to have an effect on MoM distributions but when corrected for, the changes in risk estimates were minimal and it can therefore be assumed to be of no benefit to correct for maternal height.

Other potentially more significant maternal factors are diabetes and
race. In the case of diabetes, it was shown that the AFP MoM was significantly lower than in non-diabetics. The HCG levels were not affected. These results are in agreement with other published studies and overall will have the effect that the estimated risk in diabetic women will be higher than in similar non-diabetic women. The effect of this is to increase the false positive rate in diabetics. Since it is known that there is a higher rate of birth anomalies in diabetics, this effect causing increased risk estimates can reasonably be considered unimportant because it may simply mean that diabetic women get more careful antenatal care.

The evaluation of maternal race presented in section 5.4 did not reveal any significant difference between AFP or HCG MoMs from different ethnic groups: this is not surprising because of the extremely small size of the non-caucasian groups. A larger study is needed to determine the true situation.

Overall the most important maternal factor is weight with diabetes forming a secondary factor that may need to be considered in hospitals where there is a large proportion of diabetic patients in the population. In other sites it is not unreasonable to ignore the role of diabetes because the most likely result is a false positive risk estimate. Since there are no studies of the effect of correction for the diabetic state, no estimates of the risk of false negatives are available and this means that it would be difficult to justify medico-legally, correction for diabetes in the event of being presented with a case of Down syndrome associated pregnancy where uncorrected results indicate amniocentesis and corrected results do not indicate amniocentesis.
10.1.6 What is the significance of gestation dating?

It has been demonstrated that accurate gestation dating is of vital importance and that calculation of risks using integer week derived MoMs is of dubious value due to the errors that may result from 'quantum' and 'edge' effects. Since LMP dating may be unreliable, the ideal system would be to confirm all gestation dates by ultrasound but many hospitals do not have the resources for this. A possible solution would be to calculate a range of risk estimates and only offer ultrasound to those women in whom ultrasound could make a difference. The effect of this policy has been estimated by Gordosi and Mangelli, 1993 and it has been shown that this would mean that only 14% of women would need ultrasound.

10.1.7 Which mathematical algorithm should be used?

The main question about which mathematical algorithm to use boils down to asking which algorithm is of most use to a lab that is starting a new screening program and which therefore does not have any Down syndrome data of its own. The choice of algorithm then depends partly on the choice of analytes but even after this consideration, it appears that the optimum calculation algorithm is the likelihood ratio method using the parameters of Wald et al, 1988a.

10.1.8 What can we expect from Down syndrome screening?

Estimations of the expected performance of Down syndrome screening can be derived in 3 ways: simulation; retrospective studies; and
prospective studies. Simulations are useful as a first step to see whether a particular configuration is valid but may give misleadingly optimistic or pessimistic results depending on the assumptions made in the simulation. Consequently any results derived from a simulation should be regarded with considerable scepticism.

Retrospective studies may suffer from selection bias resulting in under-representation of missed cases and a consequent over-estimation of the benefits of screening. This means that the ideal means of determining expectation is prospective trials. The study presented here is too small to give a good idea of expectations but is of a similar size to other prospective studies. The report of the Royal College of Obstetricians and Gynaecologists, 1993, quotes a series of studies of ‘double’ and ‘triple’ tests and shows that there is little difference between the two screening tests. These are however inferior to the prospective studies reported by Spencer et al, (1993b) which suggest that detection of the order of 80-90% can be expected for a 5% false positive rate using a ‘double’ combination of AFP with free β HCG.

10.1.9 What are the concerns about Down syndrome screening?

Down syndrome screening produces significant health benefits for those women whose Down syndrome fetuses are detected (whether terminated or not) but may cause excessive anxiety in women who are given a positive result that later turns out to be a false positive result. This anxiety should be minimised by ensuring that amniocentesis is available as soon as possible after reporting the ‘high’ risk result. For this reason, the RCOG report (1993) stipulates that results should not be given
on a Friday, Saturday or Sunday.

There are also health benefits for those older women who are serum tested and who avoid amniocentesis. However, it must be explained to these women that there is a chance that the serum test will miss a Down syndrome fetus to attempt to minimise the psychological sequelae if they deliver a Down syndrome baby.

Overall therefore the multi-marker Down syndrome screening test has benefits and drawbacks but the benefits appear to outweigh the drawbacks. It is not possible to quantify these in monetary terms because although the laboratory and clinical costs can be measured, the psychological effects on the women screened cannot be measured on the same scale.

The final concern about Down syndrome screening is that all Down syndrome screening programs must be large enough for the program to be assessed with respect to detection and false positive rates. It is easy to assess false positive rates but for detection rate there must be at least 10 cases of Down syndrome per year. Since there is a spontaneous 30% loss rate between the second trimester and term, this means that a minimum of 5000 tests per year should be performed for a program to be truly viable. Ideally, this figure should be greater.
10.2.1 Correlation between analytes may not be bad.

In the Wald et al, 1988a paper, it is stated that:

\[ \ldots \text{When using several variables in combination to screen for a particular} \]
\[ \text{disorder it is necessary to assess the extent of correlation among the} \]
\[ \text{variables concerned. If the two variables are perfectly correlated one adds} \]
\[ \text{nothing to the other in detecting the risk of having the disorder; if they are} \]
\[ \text{completely unrelated each provides an independent measure of risk. If} \]
\[ \text{they are partially correlated there will be some independent information.} \]

(Wald et al, 1988a)

However, computer simulations show that high degrees of correlation may actually be beneficial. Figures 10.1, 10.2 and 10.3 show contour and surface plots generated using MathCAD 4.0 software (Mathsoft, 201 Broadway, Cambridge, Massachusetts) for different assumptions about correlation between analytes having median MoMs equivalent to AFP and total HCG. From the figures it is obvious that as correlation increases, contrary to the assertion of Wald et al, discrimination between groups increases. Research to determine a well correlated markers should be performed to see whether this will offer better detection than uncorrelated markers.

Status: Paper describing the principle that correlation may be useful in preparation.
Figure 10.1.
The effect of differing relationships between analytes. $r = 0.0$
Figure 10.2.
The effect of differing relationships between analytes. $r = 0.5$
Figure 10.3.
The effect of differing relationships between analytes. $r = 0.9$
10.2.2 Correlation between analytes may be spurious.

Tables 8.5 and 8.6 show published population parameters for AFP, HCG and uE₃. It is notable that the correlation coefficients for each analyte presented by different authors vary widely in some cases from large -ve to large +ve correlations. This suggests that many of these correlations may be incorrect. A possible explanation of this could be that different authors’ populations have different degrees of gestation dating inaccuracy. To explain this: examining the correlation between AFP and HCG: As gestation age increases, the concentration of AFP increases and that of HCG decreases. Therefore if gestation age is overestimated, the AFP MoM will be falsely underestimated and the HCG MoM will be falsely overestimated. The opposite occurs if gestation age is underestimated. Thus if a large degree of dating inaccuracy is present there will be a tendency for low AFP levels to be associated with high HCG levels and vice versa. Regression analysis of such data could well lead to false correlations being found.

Status: Paper describing this effect in preparation.

10.2.3 Gestation dating may affect medians.

Error in estimated gestation age relative to ‘true’ gestation age means that a proportion of the results used to derive medians for each week will be from inappropriate age groups. Since the concentration at each week changes, this can mean that the data points selected for derivation of medians are contaminated by data from the wrong week leading to inaccurate estimates of medians.
Status: Paper describing principle and a mathematical method of deriving 'true' medians submitted.

10.2.4 Risk calculation algorithms.

The likelihood ratio risk calculation algorithm is currently based on a multivariate Gaussian assumption. This assumption is not completely met and a non-Gaussian distribution may offer a better alternative.

Status: statistical research in progress.

10.2.5 Ethnic differences.

The possibility of ethnic variation in AFP and HCG could neither be confirmed nor refuted given the small size of the data set available in Gwent. A larger sample set is necessary to evaluate this.

Status: collaborative project in progress.

10.2.6 Inter-kit variations in population parameters.

As demonstrated in section 7.1, different kits may not give comparable results. To screen effectively, it is necessary to know that different kits all give the same results.

Status: planning.
10.2.7 Within-kit variability

All assay kits are produced in batches or lots and there may be subtle differences between different lots. For the same reasons as 10.2.6, it would be useful to understand the extent of this problem.

Status: planning.

10.2.8 Within patient variation

A small study (n = 6) (Coore et al, 1993) has shown that risk estimates can be relatively constant over several weeks of gestation. This study needs to be repeated with larger numbers of patients.

Status: planning.
In 1969, Walknowska et al demonstrated that fetal lymphocytes in the maternal blood could be cultured and that chromosome analysis of these cells could be used to determine fetal sex with an excellent degree of success (19 of 21 male fetuses correctly identified). Since 1969, new techniques have been developed that allow fetal cells in maternal blood to be studied.

The polymerase chain reaction (PCR) can be used to detect fetal cell DNA without first separating it from maternal DNA. This method however, can only be used to identify fetal DNA that would not usually be present in the blood of the non-pregnant woman. For example, Y chromosome material to identify fetal sex (Lo et al, 1990); Rhesus D antigen DNA in Rhesus negative women (Lo et al, 1993); and an aberrant globin gene in a patient whose husband was a carrier for Haemoglobin Lepore-Boston disease (Camaschella et al, 1990).

To screen for Down syndrome however, it would not be possible to use a direct PCR technique because the extra chromosome 21 is not a unique entity that would not be otherwise present in the maternal blood. It is therefore necessary to separate fetal cells from maternal cells before identification of trisomic cells can proceed. Once again sorting of fetal cells is not a new technique. One of the earlier reports of a sorting technique dates from 1979 (Herzenburg et al) where a method of marking fetal cells with a rabbit anti-paternal Hla marker antibody was followed
by reaction with a goat anti-rabbit antibody tagged with a fluorescent marker. Cells sorted using this technique were tested for Y chromatin material and this allowed correct identification of 5 of 5 male fetuses. Later work using the same fluorescence activated cell sorting (FACS) method has employed antibodies to the transferrin receptor to identify nucleated erythrocytes (Bianchi et al, 1990) which when tested for its ability to diagnose male fetuses was 75% efficient. Newer FACS machines allow multiplexing of sorting characteristics and using a combination of cell size, granularity, transferrin receptor and glycophorin-A cell surface molecule, Price et al (1991) extracted cells sufficient to allow correct identification of 12 (of 12) male fetuses and 10 (of 12) female fetuses. They have also diagnosed cases of trisomy 21 and trisomy 18 but comment that further improvement in technique is needed to be certain that the sorted cells are truly fetal in origin (Elias et al, 1992).

Another group has also used the FACS technique but in conjunction with a preliminary FICOLL centrifugation step and immunomagnetic lymphocyte depletion which has been used to diagnose a case of 47, XYY (Cacheux et al, 1992). It is clear therefore that extraction of fetal cells from the maternal circulation is entirely possible and has been achieved with considerable success in reseach applications. Further development of the technique may allow its use routinely which would then open the way to true prenatal diagnosis on a maternal blood sample. This would have a significantly smaller risk than amniocentesis and could have a major effect on the future provision of ante-natal 'screening' programs.
10.3.2 FISHing for a diagnosis
(References quoted in this section are examples and not an exhaustive list)

The revolution in molecular genetics that has occurred in the last few years has led to extensive investigation of all human chromosomes. In the case of chromosome 21, markers specific to the long arm (21q) (Gardiner, 1990: Chumakov et al, 1992) and the centromere (Stewart et al, 1991) have been described and the availability of these markers has allowed more specific examination of the chromosome to determine those areas responsible for the Down syndrome phenotype. This line of approach has been followed by many investigators (McCormick et al, 1989: Stylanios et al, 1989: Van Canp et al, 1989: Rahmani et al, 1989: Tanzi et al, 1990: Korenberg, 1990: Watkins, 1990: Jankowski et al, 1990: Korenberg et al, 1992: Creté et al, 1993) which has allowed the critical region to be identified as a 1.2 megabase region around D21S55 in 21q22.2-22.3.

A spin off of this genetic research has been the ability to label chromosome specific probes with fluorescent agents. These labelled probes may then be incubated on a slide with an appropriate sample to allow in situ hybridisation to occur; hence fluorescent in situ hybridisation: 'FISH'. The possibility that this technique could be used for diagnosis was suggested as early ago as 1986 (Julien et al, 1986) and confirmed in 1988 (Lichter et al, 1988). It has now been explored by other investigators: Price et al (1991) demonstrated Y specific material in fetal cells extracted by flow cytometry; Wessman et al (1992) demonstrated similar material in FICOLL separated fetal cells. Further examples include diagnosis of trisomy 21 (Elias et al, 1992: Bryndorf et al, 1992: Evans et al, 1992) and an
It is thus obvious that FISH has the potential to be a useful prenatal diagnostic tool if sample material can be extracted. The problem with FISH is that it relies on there being sufficient cells in metaphase and interphase to give a sufficient data set for diagnosis and that it is a visual technique which unless it can be automated by image analysis may be excessively resource intensive for it to be offered to all pregnant women.

10.3.3 PCR Possibilities

The second molecular genetic technique that may eventually replace serum screening is PCR. A semi-quantitative method for identification of Down syndrome cells has already been described (Miller et al, 1992). Unfortunately however, PCR requires a purer sample of fetal cells than FISH because whereas in the FISH technique it would be possible to stain for a fetal product (e.g. HbF) and only count cells that are so marked, this is not possible in PCR because this is a test tube technique that does not involve identification of any individual cells. In the long term however, PCR offers the most attractive method for prenatal diagnosis because it has far greater potential for automation and may eventually become as simple as any other colorimetric clinical chemistry technique.


Brousseau K. (1928). *Mongolism: A study of the physical and mental


Br. Med. J. 305: 1017


real time ultrasound. Ultrasound in Med. 1: 97-104.


press.


Macri J. (1990b). Triple test: premature, counterproductive. Perspectives in Genetic Counselling. 12: 1


Diagn. 10: 575-81


Oliver C. (1891). A clinical study of the ocular symptoms found in so-called mongolian idiocy. Trans. Ophth. Soc. 6: 140


Penrose L. (1933). The relative effects of paternal and maternal age in mongolism. J. Genet. 27: 219


and between batch variability of the estimation of Down’s syndrome risk.
Proceedings of the ABD National meeting: 37


Tabor A, Larsen S, Neilsen J, Neilsen Joh, Philip J, Pilgaard B, Videbach P,


Tonks D. (1968). A quality control program for quantitative clinical


Wilmarth S. (1890). Report on the examination of 100 brains of feeble-
minded children. Alienist and Neurologist. 11: 520


APPENDIX 1: COMPUTER PROGRAM SUBROUTINES

All subroutines are written in Borland Turbo PASCAL unless stated otherwise.

A1.1 Age-Related Risk Calculation

FUNCTION AgeRisk (Age: real): real;
Var
value : real;
begin
value := power(e, ((0.286 * Age) - 16.2395));
AgeRisk := (0.999373 + value) / (0.000627 + value);
end;

This subroutine requires a subroutine to calculate powers e.g. x^3. Also the value of e must be declared as a constant. See Appendix 1, section A1.2.

A1.2 Power Calculation

FUNCTION Power (x,y: real): real;
{x must be a positive real number}
Var
value : real;
begin
value := LogTen(x) * y;
Power := AntiLogTen(value);
end;

This subroutine requires subroutines to calculate Log_{10} and 10^x. See Appendix 1, sections A1.3 and A1.4.

A1.3 Log_{10} Calculation

FUNCTION LogTen (x: real): real;
{x must be a positive real number}
begin
LogTen := LN(x) / LN(10);
end;

The function LN is provided in the PASCAL but calculates LOG_e.
A1.4 Anti-LOG\(_{10}\)

FUNCTION AntiLogTen (x: real): real;
Var
value : real;
begin
value := LN(10) * x;
AntiLogTen := EXP(value);
end;

A1.5 Weight Correction of MoMs (least squares logarithmic method)

FUNCTION Weight_Correct (MoM, Weight, A, B: real): real;
{A, B = regression parameters appropriate to the analyte being weight corrected}
begin
Weight_Correct := MoM / (AntiLogTen((A-(B * Weight)));
end;

This subroutine requires a subroutine to calculate 10\(^x\). See Appendix 1, section A1.4.

A1.6 Univariate Gaussian Distribution Calculation

FUNCTION Fx (mu, sig, x: real): real;
{Expects mu and sig (the population mean and sd), and x (Patient value)}
Var
Mahalanobis, value: real;
begin
Mahalanobis := Power(((x - mu)/sig),2);
Value := Power(e, (-0.5 * Mahalanobis));
Value := (1/(sig * SQRT(2 * Pi))) * Value;
Fx := Value;
end;

This subroutine needs values for e and defined as constants in the program.
A1.7 Bivariate Gaussian Distribution Calculation

FUNCTION Fxy (mu, sig, x : array[0..1] of real; Rho: real): real;

{Expects mu and sig (the population mean and sd), and x (Patient value) as arrays e.g. mu[0] and mu[1] contain the mean values for parameter 1 and 2 respectively. Also requires Rho which is the correlation coefficient between parameters 1 and 2}

Var
  Mahalanobis, Inter1, Inter2 : real;
  Value : array[0..1] of real;
  i : integer;

begin
  For i := 0 to 1 do
    Value[i] := (x[i] - mu[i])/sig[i];
  end;
  Mahalanobis := Power(Value[0],2);
  Mahalanobis := Mahalanobis + Power(Value[1],2);
  Mahalanobis := Mahalanobis - (Value[0] * Value[1] * 2 * Rho);
  Inter1 := 1/(1 - Power(Rho,2));
  Inter1 := Power(e, (-0.5 * Inter1 * Mahalanobis));
  Inter2 := sig[0] * sig[1] * 2 * Pi * SQRT(1 - Power(Rho,2));
  Inter2 := 1 / Inter2 * Inter1
  Fxy := Inter2;
end;

This subroutine needs values for e and Pi defined as constants in the program.

A1.8 Trivariate Gaussian Distribution Calculation

FUNCTION Fxy (mu, x : array[0..2] of real; V_matrix : array[0..2,0..2] of real; det: real) : real;

{Expects mu (the population means), and x (Patient values) as arrays. Also requires V_matrix which is the inverse of the covariance matrix derived from the population standard deviations and the correlation coefficients between parameters and det which is the determinant of the covariance matrix.}

Var
  Mahalanobis, Inter1, Inter2 : real;
  Value, Mu_Val : array[0..2] of real;
  i, j : integer;

begin
  For i := 0 to 2 do
    For j := 0 to 2 do
      Value[i] := (x[i] - mu[i])/sig[i];
    end;
    Mahalanobis := Power(Value[i],2);
    Mahalanobis := Mahalanobis + Power(Value[j],2);
    Mahalanobis := Mahalanobis - (Value[i] * Value[j] * 2 * Rho);
    Inter1 := 1/(1 - Power(Rho,2));
    Inter1 := Power(e, (-0.5 * Inter1 * Mahalanobis));
    Inter2 := sig[i] * sig[j] * 2 * Pi * SQRT(1 - Power(Rho,2));
    Inter2 := 1 / Inter2 * Inter1
    Fxy := Inter2;
  end;
end;
begin
For i := 0 to 2 do
begin
   Value[i] := x[i] - mu[i];
end;
For i := 0 to 2 do
begin
   Mu_Val[i] := 0;
   For j := 0 to 2 do
   begin
      Mu_Val[i] := Mu_Val[i] + (Value[j] * V_matrix[j,i]);
   end;
end;
For i := 0 to 2 do
begin
   Mahalanobis := Mahalanobis + (Mu_Val[i] * Value[i]);
end;
Inter := Power(e, (-0.5 * Mahalanobis));
Inter := 1/(SQRT(Power((2 * Pi),3))) * Inter * (1 / det);
Fxy := Inter;
end;

This subroutine needs values for e and defined as constants in the program.

A1.9 Determinant Calculation

These procedures are written in Acorn Archimedes BASIC V. There are a number of special commands invoked in these routines that are only available in this version of BASIC. These are:

matrix()/a : Divides all elements of the matrix by a.
matrix()*a : Multiplies all elements of the matrix by a.
matrix1().matrix2() : Performs matrix multiplication (Chapter 3, section 3.2.5.2.3)

PROCdeterminant as presented here will calculate a determinant for a 6 x 6 matrix but may be expanded further. Calculating the determinant for a 10 x 10 matrix with this subroutine takes approximately 24 hours.
PROCmatrices_Initialise must be called first because this sets up the Global arrays used by PROCdeterminant.

10   DEF PROCmatrices_Initialise
20   DIMb5(5,5),b4(4,4),b3(3,3),b2(2,2),b1(1,1)
DEF PROC determinant(a(),RETURN deter)
    IF DIM(a(),1)=1 deter=(a(0,0)*a(1,1))-(a(1,0)*a(0,1)):ENDPROC
    LOCAL C%
    deter=0
    FOR C%=0 TO DIM(a(),1)
        CASE DIM(a(),1) OF
            WHEN 6: PROC construct_minor(a(),b5(),C%))
            PROC determinant(b5(),deter)
            deter6=deter6+(deter*((-1)ˆC%)*a(0,C%))
            WHEN 5: PROC construct_minor(a(),b4(),C%))
            PROC determinant(b4(),deter)
            deter5=deter5+(deter*((-1)ˆC%)*a(0,C%))
            WHEN 4: PROC construct_minor(a(),b3(),C%))
            PROC determinant(b3(),deter)
            deter4=deter4+(deter*((-1)ˆC%)*a(0,C%))
            WHEN 3: PROC construct_minor(a(),b2(),C%))
            PROC determinant(b2(),deter)
            deter3=deter3+(deter*((-1)ˆC%)*a(0,C%))
            WHEN 2: PROC construct_minor(a(),b1(),C%))
            PROC determinant(b1(),deter)
            deter2=deter2+(deter*((-1)ˆC%)*a(0,C%))
            ENDCASE
        NEXT
        CASE DIM(a(),1) OF
            WHEN 5: deter=deter5:deter5=0
            WHEN 4: deter=deter4:deter4=0
            WHEN 3: deter=deter3:deter3=0
            WHEN 2: deter=deter2:deter2=0
            ENDCASE
        ENDPROC
        DEF PROC construct_minor(m(),RETURN mm(),C%)
        LOCAL i%,j%,j
        FOR i%=0 TO DIM(m(),1)-1
            FOR j%=0 TO DIM(m(),1)-1
                IF j%>=C% THEN j%=j%+1 ELSE j%=j%
                mm(i%,j%)=m(i%+1,j)
            NEXT
        NEXT
        ENDPROC
A1.10 Matrix Inversion

These procedures are written in Acorn Archimedes BASIC V.

500 DEF PROC inverse(toinvert(), RETURN inverted(), inter(), inter1())
510 CASE DIM(toinvert(), 1) OF
520 WHEN 1:
530 PROC determinant(toinvert(), det)
540 inverted(0, 0) = toinvert(1, 1)
550 inverted(1, 1) = toinvert(0, 0)
560 inverted(0, 1) = -1 * toinvert(0, 1)
570 inverted(1, 0) = -1 * toinvert(1, 0)
580 inverted() = inverted() / det
590 OTHERWISE
600 PROC transpose(toinvert(), RETURN transposed())
610 PROC minors matrix(inter(), inter1(), inverted())
620 PROC determinant(toinvert(), det)
630 inverted() = inverted() / det
640 ENDPROC
645
650 DEF PROC transpose(totranspose(), RETURN transposed())
660 FOR II% = 0 TO DIM(totranspose(), 1)
670 FOR JJ% = 0 TO DIM(totranspose(), 1)
680 transposed(JJ%, II%) = totranspose(II%, JJ%)
690 NEXT: NEXT
700 ENDPROC
705
710 DEF PROC minors matrix(am(), am1(), RETURN e())
720 LOCAL i%, j%, I, J
730 FOR im% = 0 TO DIM(am(), 1)
740 FOR jm% = 0 TO DIM(am(), 1)
750 FOR i% = 0 TO DIM(am(), 1) - 1
760 FOR j% = 0 TO DIM(am(), 1) - 1
770 IF i% >= im% THEN I = i% + 1 ELSE I = i%
780 IF j% >= jm% THEN J = j% + 1 ELSE J = j%
790 am1(i%, j%) = am(I, J)
800 NEXT: NEXT
810 PROC determinant(am1(), det)
820 e(im%, jm%) = ((-1) ^ (im% + jm%)) * det
830 NEXT: NEXT
840 ENDPROC
A1.11 Uniform Random Number Generator

FUNCTION Random_Number(var ix,iy,iz :integer): real;
{Original supplied by Christine Donovan}
{This procedure generates a pseudo-random variable rectangularly
distributed between 0 and 1}
{ix, iy and iz are generated seeds set to integer values between 1 and
30000 before entry. These values are set only once as they are altered
every time the function is called ensuring a random result}
{Integer arithmetic up to 30323 is required}
Var
   irand : real;
begin
   ix := 171 * (ix mod 177) - 2 * (ix div 177);
   iy := 172 * (iy mod 176) - 2 * (iy div 176);
   iz := 170 * (iz mod 178) - 2 * (iz div 178);
   If (ix < 0) then ix := ix + 30269;
   If (iy < 0) then iy := iy + 30307;
   If (iz < 0) then iz := iz + 30323;
   irand := (ix / 30269) + (iy / 30307) + (iz / 30323);
   irand := irand - trunc(irand);
end;

A1.12 Random Normal Deviate Calculation

PROCEDURE nrand (n1 : integer;
   var ix,iy,iz :integer;
   var sim : vector1);
{Original supplied by Christine Donovan}
{This procedure generates normal random variables from two
independent uniforms by the Polar Marsaglia-Bray method. Results are
supplied in the array sim}
{vector1 is defined as type = array[1..5] of real}
Var
   i,nhalf :integer;
   u,v,vv :real;
begin
   nhalf:=(n1+1)DIV 2;
   for i:=1 to nhalf do begin
      Repeat
         u := Random_Number(ix,iy,iz); {u and v are independent u(0,1),}
         v := Random_Number(ix,iy,iz); {transformed to u(-1,1)}
         u:=(2*u)-1;
      end;
end;
\[ v := (2v) - 1; \]
\[ v_{v} := (u^2 + v^2); \quad \{ u \text{ and } v \text{ are now transformed to two IID N}(0,1)'s \} \]
Until \( v_{v} \leq 1 \);
\[ v_{v} := \sqrt{(-2 \ln(v_{v})/v_{v})}; \]
\[ \text{sim}[i] := u \times v_{v}; \]
\[ \text{sim}[n1-i-1] := v \times v_{v}; \]
end;
end;

**A1.13 Cholesky Root Calculation**

PROCEDURE Cholesky (n1,p1 : integer;
var v,t : vector2;
var Cholesky_OK : Boolean);
{Original supplied by Christine Donovan}
{Calculates cholesky square root matrix (t) of covariance matrix v which
is assumed to be positive definite. The routine requires requires a lower
triangular matrix t st t*t(tr) = v}
{vector 2 is declared as a type = array[1..5,1..5] of real}
Var
\[ te, s \quad : \text{real}; \]
\[ i, i1, j, j1, k \quad : \text{integer}; \]
Done \quad : \text{boolean};
begin
Cholesky_OK := True;
Done := False;
For i:= 1 to p1 do
begin
For j:= 1 to n1 do
begin
\[ t[j,i] := 0 \]
end;
end;
\[ te := v[1,1]; \]
If \( te < 1 \times 10^{-12} \) then Cholesky_OK := False;
If Cholesky_OK then
begin
\[ t[1,1] := \sqrt{te}; \]
if n1=1 then Done := True;
If NOT Done then
begin
\[ t[2,1] := v[2,1]/t[1,1]; \]
\[ te := v[2,2]-sqr(t[2,1]); \]
If \( te < 1 \times 10^{-12} \) then Cholesky_OK := False;
If Cholesky_OK then
begin
  t[2,2]:=sqrt(te);
  If n1=2 then Done := True;
  If NOT Done then
  begin
    For i:=3 to n1 do
    begin
      t[i,1]:=v[i,1]/t[1,1];
      i1:=i-1;
      For j:=2 to i1 do
      begin
        s:=0;
        j1:=j-1;
        for k:=1 to j1 do
        begin
          s:=s+t[i,k]*t[j,k];
        end;
        t[i,j]:(v[i,j]-s)/t[j,j];
      end;
      s:=0;
      For j:=1 to i1 do
      begin
        s:=s+sqr(t[i,j]);
      end;
      te:=v[i,i]-s;
      If (te< 1E-12) then Cholesky_OK := False;
    end;
  end;
end;

A1.14 Multivariate Normal Distribution Simulation

PROCEDURE Multivariate_Normal_matrix
(No_of_parameters :integer;
  var ix,iy,iz :integer;
  mean :vector1;
  t :vector2);
This subroutine generates a random multivariate normal sample of size 'No_of_parameters' and 'mean' vector by converting the output of nrand (which generates independent normal samples distribution N(0,1)) to a correlated dataset. This is achieved by using the Cholesky root decomposition product of the covariance matrix (supplied as t).

Var
    dim, dim2: integer;
    Normal_set: vector1;
begin
    nrand(No_of_parameters, ix, iy, iz, Normal_set);
    for dim := 1 to No_of_parameters do
        begin
            x[dim] := mean[dim];
            for dim2 := 1 to No_of_parameters do
                begin
                    x[dim] := x[dim] + t[dim, dim2] * Normal_set[dim2];
                end;
        end;
APPENDIX 2: ADDRESSES OF SUPPLIERS

Abbott Diagnostics Division,
Abbott House,
Moorbridge Road,
Maidenhead,
Berkshire,
SL6 8XZ.

Bio-Rad Laboratories (UK) Ltd.,
Mayland Ave,
Hemel Hempstead,
Hertfordshire,
HP2 7TD.

CIS (UK) Ltd.,
Dowding House,
Wellington Road,
High Wycombe,
Berkshire,
HP12 3PR.

Dako Ltd.,
16 Manor Courtyard,
Hughenden Avenue,
High Wycombe,
Bucks.,
HP13 5RE.

DPC,
Glyn Rhonwy,
Llanberis,
Caernarvon,
Gwynedd,
LL55 4EL.

IDS Ltd.,
Bolton Business Park,
Tyne and Wear,
NE35 9PD.
Kodak Diagnostics Ltd.,
Mandeville House,
62 The Broadway,
Amersham,
Bucks.,
HP7 0HJ.

Merck Ltd.,
Hunter Boulevard,
Magna Park,
Lutterworth,
Leicester,
LE17 4XN.

NIBSC,
Holly Hill,
Hampstead,
London.

Oxoid USA Ltd.,
PO Box 691,
Ogdensburg,
NY 11771,
USA.

Pharmacia Wallac Ltd.,
20 Vincent Avenue,
Crownhill Business Centre,
Crownhill,
Milton Keynes,
MK8 0AB.

Sigma Diagnostics Ltd.,
Fancy Road,
Poole Dorset,
BH17 7NH.
APPENDIX 3: ABSTRACTS OF PAPERS COMPRISING THIS THESIS

Abstracts are the abstracts published as a summary of the paper except where no abstract was originally published. In these cases an abstract has been written for the purposes of this appendix. These abstracts will be identified by the note (No original abstract) at the end of the summary.

The mathematical basis of multivariate risk screening: with special refernce to screening for Down's syndrome associated pregnancy.

The underlying mathematical techniques behind the multivariate normal distribution are explained in principle and expanded in detail for up to three variables. The application of multivariate distributions to screening programs is described with particular reference to the calculation of risk ratios for the detection of Downs syndrome between 15 and 19 weeks of gestation.

In Down syndrome screening, biparietal diameter : femur length ratio should be corrected for gestational age using biparietal diameter and not biparietal diameter derived gestational age.

The biparietal diameter : femur length ratio has been suggested to be an ultrasound marker for Down syndrome but it has been demonstrated that the ratio is gestational age dependent. It is recommended that standardisation for gestation age should be performed directly via the BPD instead of converting BPD to a date expressed as integer weeks. (No original abstract).

The effect of weight correction on risk calculations for Down syndrome screening.

Recent advances in prenatal screening have lead to the possibility that the risk of Down's syndrome associated pregnancy may be assessed by blood tests for maternal serum alpha-fetoprotein, human chorionic
gonadotropin (and possibly unconjugated oestriol) taken at 15 - 18 weeks.

In neural tube defect screening correction of maternal serum alpha-fetoprotein for maternal weight has been recommended, although the precise method for weight correction is still under debate. We report an assessment of weight correction for maternal serum alpha-fetoprotein and human chorionic gonadotropin based on 1408 singleton pregnancies and for unconjugated oestriol based on 197 singleton pregnancies. We demonstrate that weight correction of maternal serum alpha-fetoprotein and human chorionic gonadotropin is statistically valid but that correction of unconjugated oestriol is not.


The effect of gestation dating method and analytical imprecision on Down syndrome risk screening are evaluated by a combination of experimental and computer simulation methods. It is concluded that gestational age must be expressed as week + day as a minimum and that expression as integer week is unsatisfactory. The examination of analytical imprecision of two parameter (AFP + HCG) screening shows that this achieves acceptable precision but that the imprecision of triple screening (AFP + HCG + uE3) appears to be marginal or unacceptable.

Ultrasonic dating of pregnancy causes significant errors in Down syndrome risk assessment that may be minimised by the use of biparietal diameter based means. Am J. Obstet. Gynecol. 166: 872-7

Objectives: Gestational dates assessed by ultrasonic measurement of fetal dimensions are usually quoted in terms of complete weeks because the uncertainty of ultrasonic measurement is approximately ±7 days. This study examines the effect of ultrasonographic dating on Down syndrome risk assessment.
Study Design: The effect of small changes in measured biparietal diameter resulting in a change in estimated gestational week and the benefits of a more precise measure of fetal gestational age (raw biparietal diameter) are examined mathematically.
Results: If maternal serum α-fetoprotein are used to assess Down syndrome risk, risks assessed with ultrasonographically determined dates
may be \(\leq 45\%\) high for fetuses with biparietal diameter at the lowest end of the size band and \(\leq 22\%\) too low with biparietal diameter at the top end of the size band. If the results for maternal serum \(\alpha\)-fetoprotein, human chorionic gonadotropin and unconjugated estriol are used, these figures become \(\leq 150\%\) and \(\leq 60\%\) respectively.

**Conclusions:** If ultrasonography is used to assess gestational age, the raw biparietal diameter and not the estimated week of gestation must be used to derive means for calculation of multiples of the mean.

6) **Reynolds T, John R. (1992).**
Comparison of assay kits for unconjugated estriol shows that expressing results as multiples of the median causes unacceptable variation in calculated risk factors for Down syndrome.
Clin. Chem. 38: 1888-93

We compared the performance of two methods for assaying unconjugated estriol in serum: the modified Amerlex third trimester RIA kit, as used in the seminal papers on unconjugated estriol in Down syndrome screening, and the new optimized Amerlex-M second trimester kit. The significant difference between the results of each assay could cause unacceptable changes in the detection rate and false positive rate of Down syndrome screening programs, especially if previously published values for estriol are used in the risk calculation. It is not possible to define new calculation parameters for every assay kit because new parameters will need to be defined every time kit changes occur, which would require a large collection of samples from Down syndrome pregnancies for standardisation. Possible solutions to this are discussed.

7) **Reynolds T, Nix B, Dunstan F, Dawson A. (1993).**
Age-specific detection and false positive rates: An aid to counseling in Down syndrome risk screening.
Obstet. Gynecol. 81: 447-50

**Objective:** To determine whether the serum screening test for Down syndrome provides equal detection efficacy for women of all ages, to improve the data available for patient counseling both before testing and afterward in the event of a positive result.

**Methods:** We examined the effect of age on Down syndrome screening by generating a set of ‘normal’ and ‘Down syndrome’ likelihood ratios by compute simulation. The expected false positive and detection rates were derived for different age groups by counting the proportion of cases in which the likelihood ratio could modify the age specific risk to be greater
than the cut off risk of one in 300 (equivalent to an incidence of 3.33 per 1000). The predictive value of a positive result was calculated using Baye's theorem.

Results: Detection rates, false positive rates and predictive values were shown to be age dependent.

Conclusions: Knowledge of the age dependency of Down syndrome screening may be useful in explaining to patients that the Down screen can only detect a proportion of cases and that a negative result does not guarantee normality. This knowledge may also be helpful in minimising psychological stress, as a positive result indicates only a small chance that the fetus will have Down syndrome.

Maternal age and recommendation for amniocentesis.
Am. J. Med. Genet. 45; 345

The veracity of lowering the maternal age limit to increase detection of aneuploidies as recommended by Kaffe and Hsu (1992) is questioned because offering screening on the basis of age alone is ineffective because acceptance of amniocentesis when offered on the basis of age alone is poor. (No original abstract).

Weight correction revisited: Does maternal height affect maternal serum AFP and HCG levels.

Maternal weight correction of AFP MoMs has long been established in NTD screening and has also been shown to have some benefit in Down syndrome screening. Since weight correction is a surrogate for a correction factor dependent on a volume of distribution within the maternal water spaces, and weight is significantly correlated with height, the relationship between height and analytes used in Down syndrome screening was investigated.

It was shown that height does significantly correlate with AFP MoMs even after weight correction but that there was no relationship between HCG and height. The effect of height on Down syndrome risks however is so minimal that it may be safely ignored.

Serum screening for Down syndrome: A joint report for the Welsh Office

This report describes a prospective study carried out at the University Hospital of Wales, Cardiff, Royal Gwent Hospital, Newport and Nevill Hall Hospital, Abergavenny from February 1990 to February 1991. During the study year, 9937 women booked for delivery. Of these 9283 were eligible for screening and 8414 accepted it. There were 296 'screen positive' results (3.5% of those tested). 17 Down syndrome cases occurred during the study year; 7 in the screen positive group; 7 in the screen negative group (although one of these had a risk of 1:303 and was offered amniocentesis with consequent prenatal diagnosis); and 3 in the unscreened group, two of which were diagnosed by amniocentesis for maternal age. (No original abstract).


This report is a summary of the report to the Welsh Office. (No original abstract).


Free β-HCG is a new analyte that has been suggested to be superior to total HCG when used in combination with alpha-fetoprotein (AFP) for Down syndrome risk screening in early pregnancy. We have evaluated this claim on 21 samples collected from Down syndrome pregnancies and 180 samples from unaffected pregnancies. The detection rates for the combination of AFP with free β-HCG or the combination of AFP with total HCG were identical (71%) but the initial screen positive rate (equivalent to false positive rate) was 7.5% for AFP + free β-HCG compared with 3.5% for AFP + total HCG screening. We conclude that the case for free β-HCG is unproven and suggest that further data be collected before free β-HCG becomes acceptable.

The statistical procedure for discriminating between a Down's syndrome or neural tube defect (NTD) fetus and a normal fetus relies to a great extent on the reporting of MSAFP, hCG and uE3 results in the form of multiples of the median (MoM). Further, threshold MoM values for MSAFP, such as 2.5 MoM, are often used to define a reference range to identify a NTD fetus. We show that a constant threshold MoM cut off for MSAFP values actually refers to different percentiles of MSAFP levels at different gestational ages and that the combining of MoM values between centres and gestational ages, such as suggested by Cuckle et al. (1987) for deriving a patient specific risk index, is highly questionable. The results presented in this paper are quite general and will apply to all situations when MoM’s are used.


Current software used for Down syndrome screening gives misleading risks when applied to abnormalities other than Down syndrome. Often the abnormality is reflected in maternal blood serum levels and is then translated into an erroneously low risk. This is a serious problem with the current screening interpretation algorithm. A modification is suggested that enables this problem to be overcome.


In Down syndrome screening the 'result' that is returned to the clinician is a risk estimate derived by modifying the maternal age related risk by a factor that depends on whether the overall pattern of analyte results is more typical of Down syndrome or more typical of normality. Determination of which group the results belong to is based on the major assumption that the population distribution of each analyte is Gaussian. However, the concentrations of the analytes used to determine group identity change with gestational age and this necessitates an initial standardisation step. Currently this is performed by deriving multiples of the median: a very poor method because it cannot compensate for
large methodological biases. This has been clearly demonstrated by comparing the assay used for the seminal papers on unconjugated oestriol with a properly optimised kit where the slope and intercept for comparison of MoMs were 2.32 and -1.3 respectively. If conversion to MoMs was effective these values should have been 1 and 0. Since no method has yet been developed to avoid using MoMs, they continue to be used and with well optimised assays they are relatively effective at standardising for gestational age. After LOG transformation of MoMs the population distributions (for AFP and HCG) are approximately Gaussian over their entire range and therefore the initial assumption is met.

The likelihood ratio method for derivation of Down syndrome risks results in some peculiarities in screening performance. Detection and false positive rates are age dependent with significantly lower rates in younger women. This is of great significance in counselling before testing and also in 2^o counselling when a 'screen positive' results is returned because the predictive value of a positive result varies between 1 - 4% (also age dependent).

A second consequence of the likelihood ratio method for deriving Down syndrome risks is the amplification of laboratory imprecision. The result for each analyte is used many times in the calculations and the number of uses increases as the number of analytes increases. Therefore as more analytes are used the imprecision increases and hence the confidence limits for the resultant risk become wider. Using typical imprecision levels from a routine Down screening program, the coefficient of variation (CV) of the risk is approximately 12% per analyte. Thus for 'double' testing the CV is 24% and for 'triple' testing it is 36%. In the early 1960's, it was determined by laboratory criteria that the maximum acceptable CV should be 20% and clinical surveys in the early 1980's conformed this limit. It therefore appears that using the likelihood method, only two analytes can be used to estimate Down syndrome risk.

Further areas where statistical analysis can be of help in screening are in identifying those patients whose results are very atypical of normal but which do not have a high risk of Down syndrome and in examining what the correct population parameters should be. Use of incorrect parameters can significantly reduce detection.