Relationships between Follicle, Oocyte and Embryo Features and the Developmental Potential of Human IVF Embryos: an investigation

Ruhi Saith

Wolfson College
Oxford

A dissertation submitted to the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Hilary Term, 1996

RADCLIFFE SCIENCE LIBRARY OXFORD
DEPOSITED
26 JUN 1997
THESIS

ABSTRACT

Relationships between Follicle, Oocyte and Embryo Features and the Developmental Potential of Human IVF Embryos: an investigation

Ruhi Saith, Wolfson College Oxford

Dissertation submitted for the degree of D.Phil. Hilary Term, 1996

This dissertation reports on an investigation into relationships between human embryo, oocyte and follicular features and the developmental potential of the associated IVF embryos on transfer. An improved understanding could yield insights into biological factors influencing the development process. Besides, identifying the features associated with a good developmental potential could assist embryologists in an informed selection of the 'best' embryos for transfer.

Previous studies with the same broad goal have suffered from a number of limitations. The present study attempts to overcome some of these by using the technique of 'decision tree' analysis developed in Machine Learning in Computer Science. The technique analyses data related to a sufficient number of specific examples in a group and discovers the general combinations of features characterising the group, presenting results in an easily comprehensible form. The feasibility of using the technique was first explored in a pilot study. The patterns of features characterising the 'good' embryos that resulted in triplet pregnancies on 3 embryo transfer or high hCG secretion *in vitro* and the 'bad' embryos that did not progress to the blastocyst stage on *in vitro* culture were identified. Given the promising nature of the results, a larger study, which forms the core of the dissertation, was conducted. Data included 53 features relating to the follicle, oocyte and embryo for batches of 3 embryos transferred that resulted in either a Take home baby (TH) outcome or a negative pregnancy test (NTH). The relationship between the 53 features and the 'TH' or 'NTH' developmental potential was investigated. The results identify, amongst others, two features related to the follicular stage of development as being important determinants of the TH potential, thereby confirming that developmental potential of an embryo is determined as early as the stage of the unfertilised oocyte. The embryo grading used by embryologists in the Oxford IVF Unit was identified as important, providing for the first time an unbiased corroboration of the selection process used in the Unit. The results refine the selection criteria and suggest two additional features, namely, follicular fluid volume and rate of cleavage as being worthy of further investigation, with the latter feature being indicated as the main component of the embryo grade characterising batches that resulted in a TH outcome. The relative unimportance of a number of features recorded in the patient notes but showing no relationship to development too is identified.

The method of analysis used here could have wide-ranging applicability to a number of areas within IVF. A demonstration of this is provided in this dissertation in an area of growing importance, namely, the identification of blastocysts of different quality.

The main study reported here can be extended to incorporate maternal and sperm features. Results obtained could form the basis for assisting embryologists in the selection of embryos with the 'best' developmental potential, with the possibility of an increase in IVF success rates.

Acknowledgements

I am indebted to a number of people but for whom neither Oxford nor this thesis would have happened. In particular, I would like to acknowledge my gratitude to:

The donor of the Felix Scholarship which opened a brave new world to me

Professor Barlow for accepting me as a Graduate Student and supporting my project

Ian Sargent for his enthusiasm and interest, supervision and contribution to the new project that became this dissertation, for support in times of despair and for being a wonderful person

Professor Donald Michie who gave most generously of his time and insight and has made a singular contribution to the pages that follow; whose indispensable good counsel, vision, wide-ranging knowledge and scholarly approach have shaped this thesis in ways so numerous and subtle that I can no longer count them

The Oxford University Computing Laboratory, particularly the Machine Learning Group for their enthusiastic collaboration and support

The embryologists and other staff in the IVF unit particularly Barbara Evans and Caroline Ross for the meticulous recording of information in patient files and for allowing me access and helping me understand these records, without which most analyses in this thesis could not have been done; Andres Lopez-Bernal for his willingness to provide IVF statistics from his database; Helen for always readily looking and finding any record however old and dusty

Anuja Dokras for teaching me the first principles of IVF and her data on the blastocyst which contributed to Chapters 4 and 6

Nick Bersinger for a fruitful collaboration on the pregnancy specific glycoprotein and for hosting me in Switzerland

A number of people on Level 3 who have always been very supportive, in particular Thurl Wilkins for her patient introduction to the laboratory, her warmth and for never saying no; Pat Yudkin for her patience with my statistical

excursions; Joan Bellinger for her willingness to help anytime, Joan Kidd for never once blowing her top when the deadline was tomorrow or worse, today afternoon; Margaret Murphy, Dick Pri scott and Lynn Clee for all their help

Karen Martin for most exacting and helpful criticisms of the thesis

Those responsible for the ORS Award, Radhakrishnan Grant, and the Oxford University grant

Those who have made Oxford unforgettable - Roma, Quike, Victor, Margaret and Richard Jayne, Pam, Awen, Alex Coren, the house in Norham Gardens and those I met there; Wolfson College

Nitin Khot, for encouraging me and spending long hours proof-reading my dreams when I was applying to Oxford; my teachers in St Joseph's - especially Mother Pasquina, Ms Merlyn, Mrs Tappan; and in Jawaharlal Nehru Medical College - especially Dr Rajashekhar; Ms Najma, Mr Aga and many others in the Islamiya English Primary School; my friends - Revathi, Chitra, Serene, Kalpana, Meenal; Bibi for her care and affection; many well-wishers in Belgaum and my family - Naaz and Tanvir

Ashwin

Rashida

My parents, whose contribution to my life is immeasurable and but for whom all this would not have been possible.

Table of Contents

1. INTRODUCTION **2**

1.1 SOURCE OF MATERIAL 7

1.2 LEGAL AND ETHICAL CONSIDERATIONS 7

1.3 AIMS OF RESEARCH AND ORGANISATION OF THE DISSERTATION 7

1.4 PUBLICATIONS 9

2. LITERATURE REVIEW **12**

2.1 ANALYSIS OF INDIVIDUAL FEATURES 13

2.1.1 Analysis of embryo feature 14

2.1.2 Analysis of oocyte and follicular features 23

2.2 JOINT ANALYSIS OF FEATURES 29

2.3 SUMMARY 35

3. BACKGROUND AND METHODOLOGY **38**

3.1 INTRODUCTION 38

3.2 ORGANISATION OF THIS CHAPTER 38

3.3 OXFORD IVF UNIT: MATERIALS 39

3.3.1 IVF drug regime 39

3.3.2 Media 41

3.3.3 Microscopes 42

3.3.4 Plastic and glassware 43

3.4 OXFORD IVF UNIT: METHODS 45

3.4.1 Oocyte retrieval 45

3.4.2 Insemination (day 0) 49

3.4.3 Monitoring fertilisation (day 1) 50

3.4.4 Monitoring early embryo development (day 2) 53

3.4.5 Criteria used to select embryos for transfer 56

3.4.6 Embryo transfer 60

3.4.7 Monitoring embryo development in vitro 60

3.5 DECISION TREES AND RULES 61

3.5.1 Use of C4.5 for decision tree analysis 63

3.5.2 Data used to construct trees and rules 65

3.5.3 Extracting feature combinations to build trees and rules 66

3.5.4 What is the accuracy of the rules? 69

3.5.5 Simplification of trees and rules 75

3.6 CONSTRUCTING TREES AND RULES IN THIS DISSERTATION 77

3.7 SUMMARY 79

4. PILOT STUDY 81

4.1 INTRODUCTION 81

4.2 ORGANISATION OF THIS CHAPTER 81

4.3 MATERIALS AND METHODS 82

4.3.1 Feature selection. 82

4.3.2 Embryo selection 83

4.3.3 Data collection and conversion 88

4.3.4 Obtaining rules 92

4.4 RESULTS 93

4.5 DISCUSSION AND CONCLUSIONS 95

4.6 SUMMARY 97

5. BATCH STUDY 99

5.1 INTRODUCTION 99

5.2 ORGANISATION OF THIS CHAPTER 101

5.3 MATERIALS AND METHODS 102

5.3.1 Feature selection 102

5.3.2 Patient selection 103

5.3.3 Batch selection and classification 105

5.3.4 Data collection and conversion 105

5.3.5 Batch average 106

5.3.6 Obtaining rules 107

5.3.7 What does 'embryo grade' capture? 110

5.4 SCOPE OF THE STUDY 111

5.5 RESULTS 114

5.5.1 Results obtained using all 53 features 115

5.5.2 Results obtained excluding embryo grade 117

5.5.3 Further analysis 119

5.6 DISCUSSION 125

5.6.1 Rules obtained using all features (Rule Set 1) 125

5.6.2 What does 'embryo grade' capture (Rule Set 2)? 129

5.6.3 Relevance to current assessment procedures 131

5.6.4 Shortcomings of the experimental design of the batch study 134

5.7 A NOTE ON USING LOGISTIC REGRESSION 136

5.7.1 Relevance to results obtained with decision trees 140

5.8 CONCLUSION 141

5.9 SUMMARY 141

6. BLASTOCYST STUDY 144

6.1 INTRODUCTION 144

6.2 ORGANISATION OF THIS CHAPTER 144

6.3 BLASTOCYST GRADING 145

6.3.1 Blastocyst grading scheme of Dokras et al (1993) 147

6.3.2 Drawbacks of the Dokras grading scheme 148

6.3.3 Alternate grading scheme using decision trees 150

6.4 MATERIALS AND METHODS 152

6.4.1 Blastocyst selection 152

6.4.2 Feature selection, collection and conversion 154

6.4.3 Obtaining rules 156

6.5 RESULTS 157

6.6 DISCUSSION AND CONCLUSIONS 158

6.7 SUMMARY 161

7. GENERAL DISCUSSION AND CONCLUSION 163

7.1 SUMMARY 163

7.1.1 Main aspects 163

7.1.2 Contributions of the dissertation 165

7.2 FUTURE WORK: EXTENSIONS TO THE MAIN IDEA 169

7.3 FUTURE WORK: OTHER APPLICATIONS 171

7.4 END NOTE 172

APPENDIX A. FEATURES USED IN BATCH STUDY 174

APPENDIX B. ASSESSMENT OF IS ACTIVITY AND SP1 LEVELS IN ECF 181

REFERENCES 208

Abbreviations

BG1	Grade 1 blastocyst
BG2	Grade 2 blastocyst
BG3	Grade 3 blastocyst
check 1	first embryo check (morning), on day of transfer
check 2	embryo check just before transfer.
Con A SLR	concanavalin Stimulated Lymphocyte Reaction
day 0	day of oocyte recovery and insemination
day 1	day of fertilisation
day 2	day of embryo transfer
ECF	embryo culture fluid
ELISA	enzyme linked immunosorbent assay
ff	follicular fluid
GnRH	gonadotrophin releasing hormones
hCG	human chorionic gonadotrophin
HFEA	human fertilisation and embryology authority
HSA	human serum albumin
ICM	inner cell mass
IS	immunosuppressive
IVF	*in vitro* fertilisation
NTH	no take home baby (negative pregnancy test)
PAF	platelet activating factor
pb	polar body
SP1	pregnancy-specific beta-1 glycoprotein
T6	Tyrode's 6
TH	take home baby (successful IVF treatment cycle)
VM	vacuolated morula
χ^2	chi-square

Chapter One. Introduction

1.1 Source Of Material 7

1.2 Legal And Ethical Considerations 7

1.3 Aims Of Research And Organisation Of This Dissertation 7

1.4 Publications 9

1. INTRODUCTION

Nearly three decades have now passed, since Louise Brown was but a twinkle in a scientist's eye. It was in 1969 that Edwards and colleagues reported on the first successful fertilisation of human *in vitro* matured oocytes with *in vitro* capacitated sperm. Although initially the progress from *in vitro* fertilisation (IVF) to the successful use of such embryos for the treatment of infertility was quite slow, IVF has now firmly moved from the research laboratory to regular clinical practice.

In its early days, IVF treatment involved the laparoscopic collection of a mature oocyte from the prospective mother and incubation with sperm from the husband. If fertilised, the single embryo was cultured up to the 6 cell to morula stage, and then transferred to the uterus. Success rates were low, with a gap as long as 6 months between the birth of the world's first and the second 'test-tube' babies (Edwards *et al*, 1980).

The use of hormones to obtain more than one oocyte and therefore more embryos, was rejected by Edwards *et al* (1980) due to the many difficulties involved in establishing the correct endocrine situation capable of sustaining a pregnancy. With the refinement of superovulation regimes however, there was a resurgence of the practice with its attendant advantages (Trounson *et al*, 1981; Hoult *et al*, 1981). It allowed the transfer of a large number of embryos to a patient, increasing the chances of a successful pregnancy. It also brought the issue of multiple

pregnancies in its wake. To minimise this serious problem, IVF Units began moves towards restricting the maximum number of embryos transferred. This was endorsed by legislation (Human Fertilisation and Embryology Authority- HFEA Code of Practice 1991, onwards) restricting the maximum number of embryos transferred to 3 in the UK. Maintaining high success rates, despite the transfer of fewer embryos however, requires the embryologists to carefully select the 'best' embryos of those available for transfer. This has led to a number of investigations into methods for identifying embryos with the highest potential of resulting in a pregnancy.

These investigations have been fuelled by technological advances such as superovulation, ultrasound, vaginal ultrasound guided oocyte aspiration techniques, new assays for measuring hormones in maternal serum as well as embryo derived products in culture fluid, and high resolution microscopy - all of which have greatly improved IVF practice and also resulted in the generation of large quantities of data. This data relates to the embryos transferred, the oocytes of origin, the ultrasound appearance of the follicles, characteristics of the maternal milieu (e.g. endometrial thickness, serum hormone levels, etc.), assessment of sperm morphology and functions and the actual outcome of any transfer. The existence of meticulous records has enabled researchers to investigate amongst other things, orderly procedures for assessing the potential of an embryo to develop and progress to a pregnancy.

A number of questions continue to beg investigation. What is it that determines whether an embryo will progress to a pregnancy? What roles are played by the gametes of origin, by the follicular characteristics, by maternal factors and by the embryo or oocyte morphology? This dissertation reports an investigation into the relationships between *features* of the embryo, oocyte and associated follicle and the *developmental potential* of transferred human IVF embryos. *Feature* refers to the morphological, biochemical and other observable or measurable characteristics defining an embryo, oocyte or follicle. *Developmental potential* refers to the potential of embryos to develop on transfer. In this dissertation when referring to the outcome of transfer, the term pregnancy is used to indicate a positive pregnancy test. This includes all patients with a positive test irrespective of whether they only had a biochemical pregnancy or went on to a term or near term pregnancy. If the outcome of transfer is referred to as clinical pregnancy, it indicates a diagnosis made by the visualisation of the gestational sac and or the foetal heart on ultrasound scan (usually 6 weeks after transfer). The term Take Home baby indicates a progression of the pregnancy to term or near-term with the delivery of a live baby. In case of spare IVF embryos cultured *in vitro*, development is assessed based on the development of the embryo to the blastocyst stage. Plate A lists some of the follicle, oocyte and embryo features that are recordable at different stages of the IVF process,

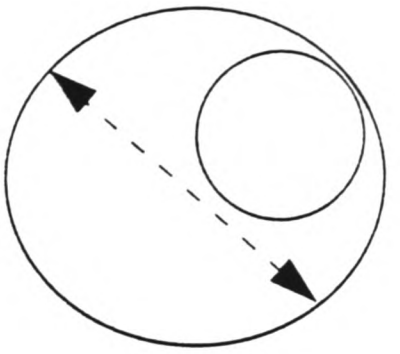
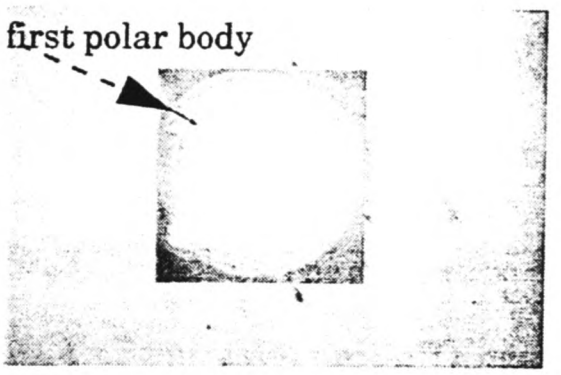
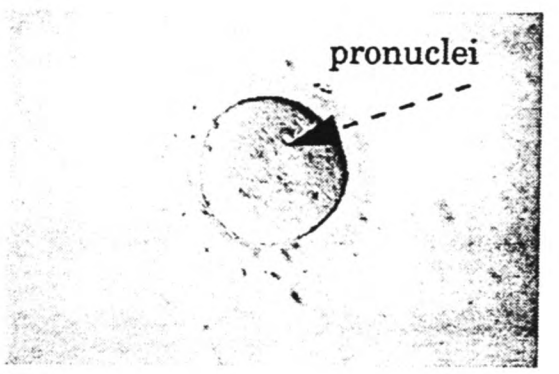
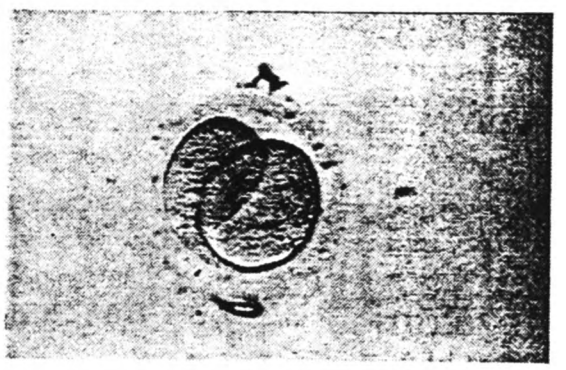
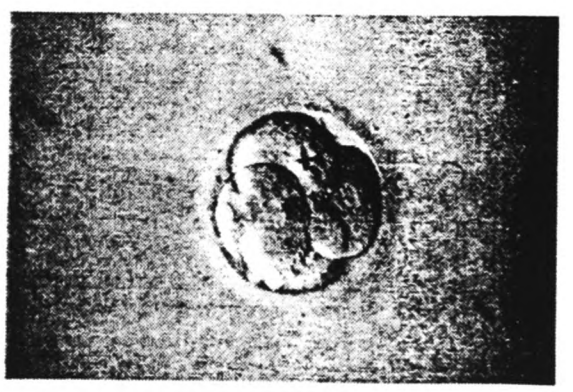
Stage of development		examples of features observed
follicle containing oocyte		diameter (by ultrasound) fluid volume levels of follicular fluid substances
unfertilised oocyte (day 0)		cumulus appearance corona appearance darkness, granularity of ooplasm
fertilised oocyte (day 1)		shape, thickness of zona pellucida darkness, granularity of cytoplasm
cleavage stage embryo (day 2)		cell number regularity of cells fragmentation shape, thickness of zona pellucida darkness, granularity of cytoplasm
cleavage stage embryo (day 2)		levels of embryo-derived products in culture fluid

Plate A. Stages of development observed during IVF treatment and examples of features that can be recorded

beginning with the ultrasound assessment of the Graafian follicle to the transfer of cleavage stage embryos.

The investigation reported in this dissertation has been motivated by a number of factors.. A clear understanding of the relationships between the features characterising the embryo, oocyte and associated follicle and the developmental potential of the embryo can yield insights into biological factors influencing the development process and the causes for infertility or failure of IVF treatment. It will add to the stock of existing scientific facts concerning the relative importance and unimportance of various combinations of features in influencing further embryo development. The added knowledge could aid clinical practice, specifically through corroborating, refining, extending and simplifying existing embryo-selection criteria.

The literature bears testimony to a number of studies with the same broad goal. These studies and their limitations are reviewed in detail in Chapter 2. Briefly, most studies restrict their research to embryo features only. The relationship between individual features and the developmental potential of embryos has been investigated, making little allowance for the influence of features related to the other stages (viz. oocyte, sperm, follicle, maternal factors) and the interaction between features. The few studies that investigate multiple features and take such interactions into account present the results as complex mathematical equations and assume linear relationships between features and development potential of embryos.

What makes the study reported in this dissertation different is that it investigates embryo, oocyte and associated follicular features, takes into account the influence of features on one another and at the same time, expresses the relationships found in the form of simple, easily understandable statements, making no assumptions of linearity. To achieve this, it uses the technique of 'decision trees' refined by research into 'Machine Learning' in Computer Science.

Data used in this research relates to embryos of known developmental outcome. This includes IVF embryos that have been transferred and the outcome of transfer (not pregnant or pregnant with the delivery of a Take Home baby) is known. Also included are IVF embryos whose development up to the blastocyst stage has been assessed by *in vitro* culture. Although an investigation of the relationship between development and maternal, sperm, follicular, oocyte as well as embryo features is important, this study is restricted to the last three groups. Features of the embryo, the oocyte of origin and associated follicle can be easily ascribed to individual embryos. Restricting the study to such features permits a simple experimental design. Inclusion of maternal and sperm features in the analysis could provide information that would assist the selection of the 'best' embryos within the *context* of the maternal environment and sperm sample used. This however is a large scale investigation that could be performed as an extension to work presented in the dissertation.

1.1 Source of material

The IVF Unit at the John Radcliffe Hospital in Oxford was the source of the data and embryos used in this dissertation.

All data related to transferred embryos was extracted from the files of patients treated in the IVF Unit. The embryos cultured *in vitro* for research purposes were spare embryos donated for research by patients being treated at the Clinic.

1.2 Legal and ethical considerations

The embryo research undertaken in this dissertation was approved by the Central Oxford Research Ethics Committee and the HFEA. Spare embryos were cultured up to a maximum period of 14 days, in accordance with the HFEA Act passed by the British Parliament in June 1991.

1.3 Aims of research and organisation of this dissertation

The principal aim of this research is to investigate the relationships between embryo, oocyte and associated follicular features and the developmental potential of human IVF embryos. Towards this end, the technique of decision tree analysis developed in Machine Learning in Computer Science has been used to discover such relationships. An improved understanding, in turn, could allow for a more informed

selection of the 'best' embryos for transfer during IVF treatment with a potential to increase IVF success rates.

The remaining chapters in this dissertation have been organised in the manner described below. The embryo, oocyte and follicular features are referred to hence forth in the dissertation as embryo-related features.:

Chapter 2. Past research investigating the relationship between embryo-related features and human embryo development is reviewed and its limitations discussed. The concluding paragraph introduces the concept of decision tree analysis.

Chapter 3. This chapter is concerned with the materials and methodology used in the dissertation. It is divided into 2 main parts. The first gives a description of the IVF treatment procedure followed at the Oxford IVF Unit with particular emphasis on the embryo-related features observed and recorded in clinical notes by the embryologists. The second part explains the decision tree analysis technique

Chapter 4. A pilot study exploring the feasibility of using decision trees to investigate relationships in IVF data is presented. Data related to transferred as well as *in vitro* cultured embryos are included.

Chapter 5. This chapter forms the core of the dissertation. It contains a large scale study exploring the relationship between embryo-related features and developmental potential. Data restricted to embryos transferred to patients in the IVF Unit is analysed.

Chapter 6. The application of ideas developed in this dissertation to other areas of IVF is contemplated. As an illustration, the use of such methods for grading blastocysts of differing quality is demonstrated. This method of grading is compared with an existing method outlined in Appendix B.

Chapter 7. The main points from the dissertation are summarised, contributions discussed and future research directions outlined.

Appendix A. The embryo-related features used in the main study in Chapter 5 are summarised.

Appendix B. A morphological grading scheme for blastocysts based on an approach largely different from that used in Chapter 6 is outlined. The correlation of these blastocyst grades with the levels of embryo derived pregnancy-specific beta-1 glycoprotein (SP1) and immunosuppressive (IS) activity in embryo culture fluid (ECF) is explored.

A brief note is required on the role of Appendix B in this thesis. Experiments in Appendix B were the first to be conducted at the start of the D.Phil. research. This continued an ongoing project in the Oxford IVF Unit. A morphology based scheme of grading blastocysts into 3 grades of differing quality had been developed. Support for the grading was being sought by measuring objective criteria (e.g. levels of embryo-derived products in the culture fluid of blastocysts) and investigating their correlation with pre-defined blastocyst grades. Here levels of SP1 and IS activity were investigated and are reported in Appendix B. This approach however had an inherent problem. For example, if levels of SP1 were

shown to correlate with blastocyst grades, they supported the grading scheme. If however, no correlation was identified, it was not clear whether it was the grading scheme that was at fault or whether the substance being measured was a poor reflector of blastocyst quality. A better approach to developing a reliable grading scheme would therefore be to use levels of a known indicator of quality, e.g. human chorionic gonadotrophin (hCG) to divide blastocysts into 'high secretors' and 'low secretors. Decision tree techniques mentioned earlier could then be used to identify patterns of morphological features characterising the blastocysts of differing quality in each of the groups. Results of applying these techniques to data at the blastocyst stage are reported here in Chapter 6. The application of these techniques was then extended to the early embryo stage which forms the core of this thesis.

1.4 Publications

The core work in this dissertation evolved from preliminary investigations that were first published as a conference abstract (Saith *et al*, 1994). A review of literature in the field was invited for publication by the Journal of Assisted Reproduction reviews (Saith R.R. and Sargent I.L., 1995).

Part of the work in Appendix B, related to the secretion of SP1 has been accepted for publication by the Journal Human Reproduction (Saith *et al*, 1996) and work related to IS activity has been presented and published as a conference abstract (Saith *et al*, 1993).

Chapter Two. Literature Review

2.1 Analysis Of Individual Features 13

2.1.1 Analysis Of Embryo Features 14

2.1.2 Analysis Of Oocyte And Follicular Features 23

2.2 Joint Analysis Of Features 29

2.3 Summary 35

2. LITERATURE REVIEW

This chapter reviews the studies that have investigated the relationships between various features related to human embryos obtained by IVF and the developmental potential of the embryo. In keeping with the scope of this dissertation, features that can be traced to individual embryos (e.g. embryo cell number or size of follicle of oocyte of origin), and define the embryo, oocyte or associated follicle are included. Although maternal and sperm features influence embryo development, they have not been included in the investigations in this dissertation and are therefore not considered in the review. Further, the 'raw material' for this dissertation was obtained from non-invasive observations of features recorded by embryologists in IVF patient notes before the transfer of selected embryos to patients. The contents of this chapter are therefore only concerned with studies using features that can be determined non-invasively.

Reports in the literature can be divided into two main groups depending on their approach. Most investigate the relationship between any one feature and the developmental potential (termed here as 'univariate' studies). These studies are reviewed first, in Section 2.1. Some other studies however, have investigated the joint relationship of a number of features with developmental potential (termed here as 'multivariate' studies). These are reviewed in Section 2.2. The different sections and their concerns are depicted in Figure 2.1.

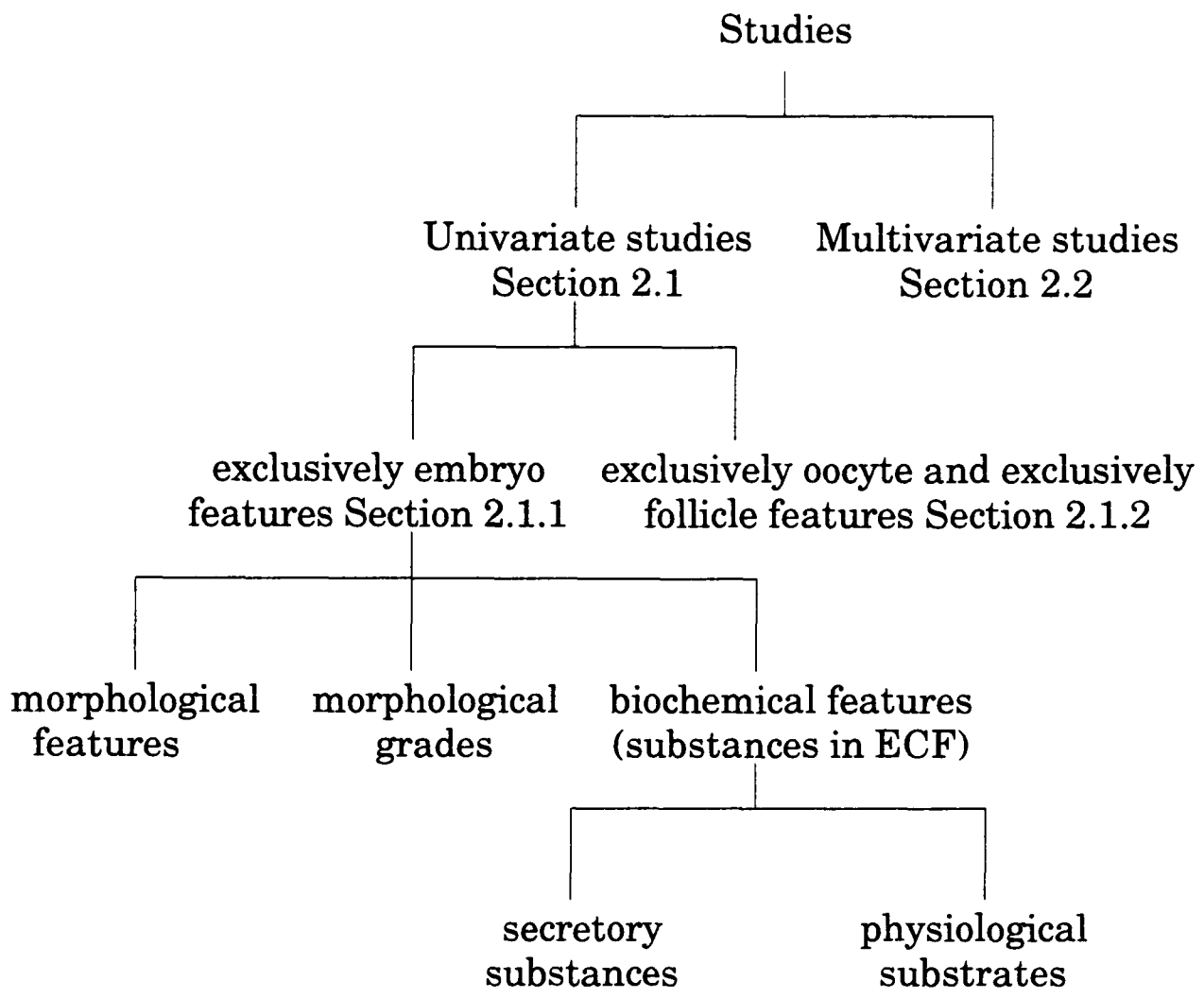


Figure 2.1 Outline of the headings under which the literature is discussed

2.1 Analysis of individual features

Literature reviewed here investigates the relationship between individual embryo-related features and development. To recapitulate, the term embryo-related encapsulates embryo, oocyte and associated follicular features. Reports concerned with embryo features are reviewed first in Section 2.1.1 followed by those that examine features related to the oocyte or the associated follicle in Section 2.1.2

2.1.1 Analysis of embryo features

Most research into assessing the relationship between embryo-related features and the developmental potential has concentrated exclusively on the embryo. The features investigated fall in 3 main groups, namely: a) morphological features; b) morphological grades and c) biochemical features, which mainly include substances measured in the ECF

A note of caution applies here. All research discussed in this section is concerned with assessing embryo features before transfer on day 2 - 3 post-fertilisation, i.e. usually before or around the 8 cell stage. Since the embryonic genome is activated at the 4-8 cell stage (Braude, *et al*, 1988), the criticism has been made that any assessment carried out before the 8 cell stage, does not exclusively relate to the embryonic genome. It is important to be aware that a proportion of the phenotype at this early stage is encoded by the maternal genome. Embryo quality however, does depend on the quality of the gametes of origin. Any relationship between the embryo features which may be coded by the maternal genome and developmental potential cannot therefore be disregarded.

Relationship between morphological features of the embryo and development

A large number of studies have sought to determine the correlation of individual morphological features to pregnancy outcome. The results reported are generally in conflict. Cohen *et al* (1989) studied the relationship between each of 11 morphological features of the embryo

(recorded using video-cinematography) with its potential to establish pregnancy. In patients with single embryo transfer, the outcome of transfer of each embryo was known. In patients with multiple embryo transfers however, to simplify the analysis, the embryo with the lowest number of abnormal morphologic characteristics was chosen to represent the batch of transferred embryos. The authors reported that an increased variation of zona pellucida thickness and a lesser degree of fragmentation correlated well with an increased incidence of clinical pregnancy (Cohen *et al*, 1989). Similarly Staessen *et al* (1992) and Dor *et al* (1986) reported that unfragmented embryos have a greater chance of resulting in pregnancy, while Claman *et al* (1987) maintain that the presence of fragments does not affect further embryo development. Conflicting reports have also been published with regard to the impact of the rate of cleavage on pregnancy. Most groups advocate the transfer of embryos that show a faster rate of cleavage (Staessen *et al*, 1992; Claman *et al*, 1987; Edwards *et al*, 1984; Trounson, 1983; Mohr *et al*, 1983). Cummins *et al* (1986) as well as Testart (1986) however, caution against the transfer of embryos that cleave too rapidly. There is some evidence to support the theory that rapid cleavage could reflect an abnormality. For example, triploid embryos generally cleave faster because of their ability to divide directly into 3 cells and then 6 cells, unlike the 2 and 4 cells usually observed for diploid cells (Plachot, 1989). In contrast, Dor *et al* (1986) suggest that the rate of cleavage has no bearing on pregnancy. One possible reason for these apparently contradictory findings is that the

research reported ignores interdependency amongst features. In the multifactorial biological systems involved in implantation, it seems extremely likely that each feature would have a bearing on, and in turn, be dependent on the others. For example, Feldberg *et al* (1990) showed that in women aged more than 40, about 33% of non-fragmented embryos achieved clinical pregnancies as compared to only 11% for fragmented embryos. Fragmentation however, had no bearing on clinical pregnancy success in younger women. A difference in the maternal or embryo-related features associated with the feature being analysed could thus be responsible for the difference in findings stated earlier.

The features found by such analyses to show a correlation with pregnancy potential have often been used to assist embryologists to assess and select embryos for transfer in IVF.

Morphology based embryo grades

Some studies have grouped embryos into grades based on a number of morphological features. The approach is as follows. Embryos are first categorised into grades of differing quality (for example: good, fair, poor) which are pre-decided based on combinations of various defining characteristics. Once the method of grading is decided, the relationship between the embryo grade (rather than the component features of the grade) and the outcome of transfer is investigated.

For example, a simple grading scheme based on 3 morphological features is shown in Table 2.1.

Table 2.1 An example of a simple morphological grading scheme for embryos

Embryo grade	Blastomere regularity	Fragmentation	Cytoplasm
1 (good or best)	regular	no fragments	clear
2 (fair)	slightly irregular	few fragments	granular
3 (poor)	very irregular	many fragments	very granular

Patterns like these are typically obtained from the experience that embryologists have had with transferring embryos with particular combinations of morphologic characteristics. Information from studies showing relationships between individual features and development is also often taken into account when formulating the grades.

Simple grading systems as in Table 2.1 have been reported by a number of groups (Staessen *et al*, 1992; Scott *et al*, 1991; Erenus *et al*, 1991; Grillo *et al*, 1991; Hill *et al*, 1989; Veeck, 1988; Plachot *et al*, 1987a; Dawson *et al*, 1987 and Yovich, 1985). More sophisticated methods that develop an embryo quality score by taking into account the rate of cleavage can be found in reports by Cummins *et al*, 1986; Steer *et al*, 1992 and Puissant *et al*, 1987.

In case of single embryo transfers, investigating the relationship between the grade of the embryo transferred and the transfer outcome is relatively simple. In case of multiple embryo transfers where the outcome of each embryo cannot be traced, some studies use the embryo with the 'best' (as defined by the authors) grade to represent the batch of transferred

embryos (e.g. Scott *et al*, 1991). Some others only consider batches in which all embryos have the same grade (e.g. Staessen *et al*, 1992). Results obtained by most groups provide evidence of a relationship between transfer outcome and embryo grades. For example, see Table 2.2.

Table 2.2 Pregnancy rates for different embryo grades in Scott *et al* (1991)

Embryo grade (grade of best embryo transferred)	Pregnancy rate
1 (best quality)	41%
2	30%
3	13%
4	22%
5 (poorest quality)	14%

Such proven morphological grades are used by many IVF Units to select embryos as it is a simple, quick, non-invasive procedure. The disadvantage is that there is no room for identifying the combinations of features that might show a higher correlation with developmental outcome than those pre-decided and allotted to the grades. In addition, if new features become available, it is impossible to incorporate them into the grading. A completely new grading scheme would have to be devised to include such features.

Criticism of the relationship between morphological appearance, either individual features or grades, and development, is often based on anecdotal accounts of apparently morphologically abnormal embryos that have resulted in pregnancy. While the studies discussed above indicate

that a particular feature or grade shows a high correlation with pregnancy, this by no means rules out the possibility that some embryos classified as poor could have a good developmental potential, albeit rarely. Such criticism therefore is unnecessary as the researchers themselves do not propose any absolute correlations. Another common criticism relates to the argument that the chromosomal anomalies found in 23-29% embryos (Plachot *et al*, 1987b and Plachot, 1988b) are not reflected in the morphological appearance. Visualisation of the pronuclei 15.5-18h after fertilisation can however eliminate most of the abnormally fertilised or parthenogenically activated zygotes. The broad consensus is that while there are no specific morphological pointers to genetic abnormalities, embryos classified as being of the poorest quality with much fragmentation do have a higher than average rate of chromosomal anomalies (Plachot *et al*, 1988a and b; Edirisinghe *et al*, 1992; Ma *et al*, 1990; Michaeli *et al*, 1990; Winston *et al*, 1991 and Pellestor *et al*, 1994) and consequently poor developmental potential. Morphological assessment and its relationship with developmental potential cannot therefore be entirely disregarded.

Relationship between biochemical features and development

This category mainly comprises two groups. Firstly, substances measured in the ECF i.e. embryo derived products, the values of which are taken to indicate the secretory capacity and well-being of an embryo. Some examples are: platelet activating factor (PAF), immunosuppressive (IS)

factor(s), lactate and pregnancy specific beta-1 glycoprotein (SP1). Secondly, the uptake of physiological nutrients from the ECF, indicative of the metabolic activity of the embryo, has also been measured.

Assessment and correlation of substances in ECF with developmental potential poses no problem with single embryo transfers where the outcome of transfer of each embryo is known. In case of multiple transfers, the approach often is to compare the average level of the substance in the ECF of all embryos transferred to patients who became pregnant against that of embryos transferred to patients who did not become pregnant.

1. **Secretory substances:** Using a bioassay, O'Neill and Saunders (1984) reported on the presence of an embryo-derived platelet activating factor (PAF) in the ECF of human embryos. PAF was found to be responsible for the peripheral blood thrombocytopaenia associated with pregnancy. Investigating the relationship between levels of embryo-derived PAF and development of the embryo, Collier *et al* (1990) reported on the presence of significantly higher levels of PAF in the medium of embryos resulting in pregnancy after transfer than those that did not. However, this finding was contradicted by Punjabi *et al* (1990) who found no correlation between PAF levels and successful pregnancy.

As with PAF, the findings of investigations into the secretion of immunosuppressive (IS) factor(s) by embryos are controversial. Research into embryo derived immunosuppressive activity as a possible

means of escaping the maternal immunological response led to the identification of suppressor factor(s) in the ECF (Daya *et al*, 1986, Bose, 1989). A number of studies report on a correlation between such suppressor activity and the potential of the embryos to progress to a pregnancy on transfer (Daya *et al*, 1986; Clark *et al*, 1989; Bose *et al*, 1990; Sheth *et al*, 1991 and Jones *et al*, 1992) although these claims have been strongly rejected by Armstrong *et al* (1989) and Armstrong and Chaouat (1993).

Lactate is a product of glucose metabolism of the embryo and levels secreted give an indication of metabolic activity which in turn reflects embryo quality. Measuring levels of lactate in the ECF of *in vitro* cultured embryos, Gott *et al* (1990) showed that embryos which subsequently arrested produced lower levels of lactate from day 2.5 onwards than those that went on to form blastocysts. The difference in lactate levels between embryos of differing developmental potential was however small, due to the low rate of glucose metabolism at the early stages of development.

2. **Uptake measurements:** These are restricted to those substances in culture media that are defined, quantifiable and serve as physiological substrates for the embryo. Hardy *et al* (1989) and Gott *et al* (1990) investigated the uptake of pyruvate which is the major energy substrate utilised by the human embryo at early stages. They showed that the pyruvate uptake on day 2.5 by embryos that later arrested was

significantly lower than that of embryos that went on to form blastocysts. Investigating single embryo transfers, Turner *et al* (1994) showed that embryos with a 'good' morphology and intermediate levels of pyruvate uptake were more likely to result in a pregnancy than those with very high or very low pyruvate uptake levels or those with a 'poor' morphology. Conaghan *et al* (1991) reported that pyruvate uptake by embryos that implanted on transfer was 18% higher than that of embryos that failed to establish a pregnancy. In contrast, in a larger study, the same authors concluded that the mean uptake of pyruvate was significantly *lower* for embryos that implanted than for those that failed to implant (Conaghan *et al*, 1993). Apart from these apparently contradictory findings, the difference in uptake between embryos of differing developmental potential on day 2 - day 3 was very small with considerable overlap of the ranges.

The findings of the above investigations into secretory substances as well as substrates, are inconclusive. Any differences in the levels of substances measured in ECF of embryos with differing developmental potential are small. The secretion or uptake of substances at this early stage, before transfer of IVF embryos on day 2 or 3, does not appear to be sufficiently defined to clearly reflect differences in developmental potential. While the clinical use of such biochemical markers to select embryos for transfer has the advantage of being non-invasive and objective, the drawbacks described above preclude their use as reliable markers. Furthermore, in many cases, the completion of the assay takes 3-4 days and the results are

not available on the day of transfer (e.g. PAF, IS activity). Those markers which can be assayed more rapidly do not show strong correlations (lactate) and can only be used together with other selection criteria.

2.1.2 Analysis of oocyte and follicular features

The quality of the oocyte from which the embryo originates has a strong bearing on its developmental potential. A number of studies have investigated the relationships between the morphological features of oocytes, secretory substances in the follicular fluid, follicle size (or follicular fluid volume) and oocyte quality or maturity. Relatively few studies, however, have sought to determine the extent of the relationship between these oocyte-related (i.e oocyte and follicle) features and the development of associated embryos, and these are discussed here.

Investigating the relationship between the maturity of the oocyte and the developmental potential of the resulting embryo, Veeck classified oocytes into three groups of varying maturity. For simplification of the study in case of multiple embryo transfers, only batches where all embryos originated from oocytes of a similar maturity status were considered. The maturity of ocytes was assessed by features observed through a high-resolution inverted microscope (Veeck, 1988). The three groups, in increasing order of maturity, were described as shown in Table 2.3.

Table 2.3 Oocyte grading scheme of Veeck (1988)

Oocyte grade	Polar body	Germinal vesicle
Prophase I	absent	visible
Metaphase I	absent	absent
Metaphase II	1 polar body visible	absent

The author reported that embryos resulting from oocytes classified as Prophase I (immature) had an extremely low incidence of pregnancy. Further, oocytes with more advanced maturity showed greater pregnancy loss. While this method was successful in predicting oocyte potential to some extent, visualisation of the polar body and germinal vesicle through the layers of cumulus and corona cells surrounding the oocyte is difficult and time consuming and the inevitable cooling of the oocyte may be detrimental. A simpler, indirect method for identifying the maturity status of the oocyte based on the appearance of the granulosa cells, cumulus and corona radiata and the ooplasm was proposed by Zenzes *et al* (1985). They too classified oocytes into 3 groups - immature, intermediate and mature and found that all pregnancies resulted from the transfer of embryos originating from mature oocytes. Both the study done by Veeck and that by Zenzes *et al*, have shown that while immature oocytes are capable of undergoing fertilisation and cleavage, they may not have a very good developmental potential. Therefore even if an embryo appears morphologically normal and has a good rate of cleavage, if the oocyte of origin is immature, it might not result in a pregnancy.

Laufer *et al* (1985) carried out an indirect assessment of oocyte maturity, measuring the products secreted by the oocyte-cumulus-corona complex. They reported that the cumulus-corona complexes of oocytes that were associated with batches of embryos resulting in a pregnancy on transfer showed a difference in steroid secretion from those batches that did not result in pregnancy.

Oocyte maturity is also reflected by the size of the follicle measured by ultrasound at the time of oocyte recovery, or the volume of the aspirated follicular fluid. Of studies investigating the correlation between follicle size (or follicular fluid volume) and pregnancy success, Kerin *et al* (1983) report no correlation between follicular fluid volume and the outcome of the associated oocyte. In contrast, most groups report on the presence of a correlation of one form or another. Lopata in 1982 showed that follicles classified as being small (<20mm diameter) or large (>20mm diameter) showed no significant difference in fertilisation or cleavage of associated oocytes. Pregnancies however resulted only from the larger follicles, more specifically from the 20-26mm diameter range. Investigating single embryo transfers, Nayudu *et al* (1987) also showed that a follicular fluid volume between 2.5-6.5ml was associated with 75% of the normal pregnancies (i.e. term or near term infant). Interestingly, they also found that the embryos associated with a majority of large follicles (>6.5ml) either produced an abnormal pregnancy (biochemical, blighted ovum and first trimester abortions) or no pregnancy. The authors speculate that this could be due to the early recruitment of these follicles during the

stimulation regime. These oocytes may be aged and therefore more prone to defects of the meiotic apparatus (Jongbloet and Zwets, 1976). On the other hand a report in 1985 by Simonetti, Veeck and Jones, investigating single embryo transfers (or transfer of 2 embryos resulting in twins) indicates that the rate of spontaneous abortion was very high when the conceptus was derived from smaller follicles (<2.0ml of follicular fluid aspirated)

Each oocyte is contained within a separate follicle bathed by follicular fluid. Levels of substances within the follicular fluid reflect the maturity of the follicle and associated oocyte and, in turn, the developmental potential of the oocyte. Findings regarding the correlation between constituents of the follicular fluid of individual follicles and developmental potential of associated embryos on transfer are however conflicting. A number of approaches have been followed. Some studies have investigated only those patients where the outcome of transfer of each embryo is known (e.g. single embryo transfer or two embryos transferred resulting in twins as in Nayudu *et al*, 1987, Simonetti *et al*, 1985, Carson *et al*, 1982). In multiple embryo transfers where this was not possible, the average level of substances in follicular fluid associated with all embryos in transferred batches that resulted in pregnancy was compared with those that did not (e.g. in Basuray *et al*, 1988; Cornwallis *et al*, 1990). Some other studies have considered the levels of substances present in all the follicles of patients who resulted in a pregnancy as compared to the

level measured in all the follicles of patients who did not result in a pregnancy.

High levels of oestradiol have been reported to be associated with the occurrence of pregnancy by Carson *et al* (1982). Most groups however disagree (Kerin *et al*, 1983; Testart *et al*, 1983; Hillier *et al*, 1985; Chabab *et al*, 1986; Franchimont *et al*, 1989; Smith *et al*, 1991; Andersen, 1990 and 1991). Similarly while Franchimont *et al* (1989) report on the presence of low levels of progesterone and Basuray *et al* (1988) support the correlation between high levels of progesterone in the follicular fluid, and the positive pregnancy outcome of associated embryos, other groups reported a lack of any correlation (Carson *et al*, 1982; Kerin *et al*, 1983; Testart *et al*, 1983; Hillier *et al*, 1985; Chabab *et al*, 1986; Smith *et al*, 1991 and Andersen, 1991). Likewise, contradictory opinions exists with regard to levels of follicular fluid androgens, with Chabab *et al* (1986) reporting the presence of lower levels in follicular fluids of oocytes associated with pregnancy while other groups reject the finding (Carson *et al*, 1982; Testart *et al*, 1983; Hillier *et al*, 1985 and Franchimont *et al*, 1989). The apparent contradictions can be partially attributed to differences in the stimulation regimes used by the groups.

Studies of other follicular fluid components have been less controversial. Tarlatzis *et al* (1985) found significantly lower levels of cyclic Adenosine Monophosphate in the follicular fluid associated with oocytes that resulted in pregnancy on transfer. They suggest that follicular fluid levels of cyclic

Adenosine Monophosphate may be a marker of optimal oocyte development. Cornwallis *et al* (1990) measured levels of renin in the follicular fluid as the ovarian renin-angiotensin system could be involved in the neovascularisation of the follicle, thus aiding the maturation of the oocyte. Comparing follicular fluid from patients who did not result in a pregnancy with those who had a normal term pregnancy they showed that significantly higher levels were associated with a successful outcome. Measuring follicular fluid from follicles that resulted in normal term pregnancies after single embryo transfer, Kreiner *et al* (1987) identified an optimum range of follicular fluid estradiol and estradiol/progesterone values. This range appears to define a stage of maturation compatible with successful pregnancies. Similarly, an optimum range of prostaglandins E_2:$F_{2\alpha}$ has been reported (Smith *et al*, 1991).

Oocyte and follicle features have not been used as extensively as embryo features to assist in the selection of embryos for transfer in IVF treatment although in some cases, oocyte maturity and follicle size have been used in conjunction with embryo features. The thresholds or ranges of values for embryo features have however been determined independently of those for oocyte and follicular features. This is unfortunate, as the possible gains of using a combination of embryo, follicle and oocyte features to assist selection could be diminished if the relationship of these multiple features jointly (explained in multivariate studies below) with development has not been taken into account.

All of the above studies have followed the univariate approach represented by Figure 2.2

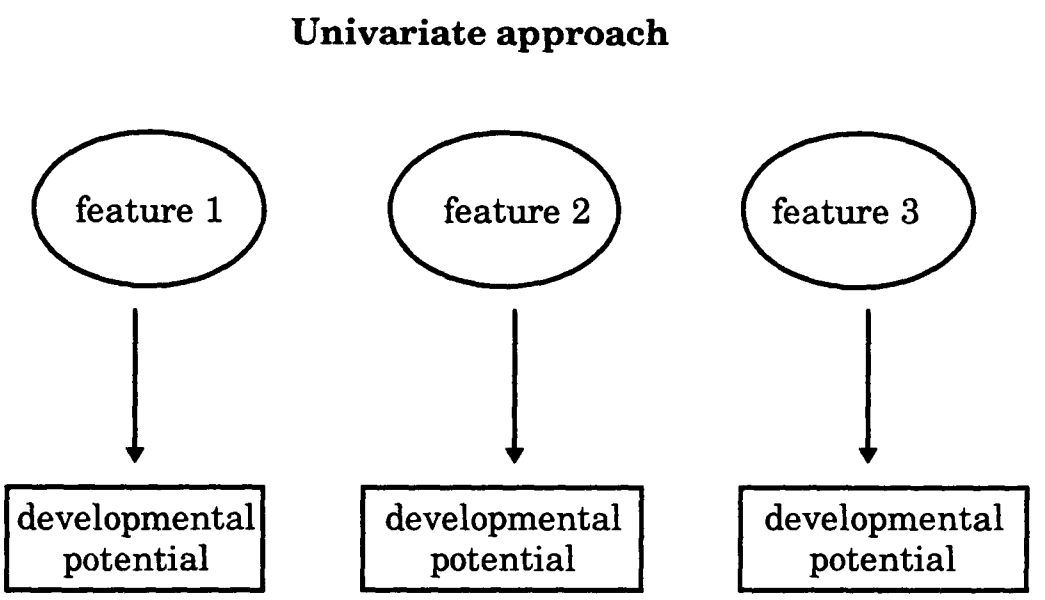


Figure 2.2 Univariate approach to analysis

If however, the influence on developmental potential of a large number of embryo, follicle and oocyte features could be investigated simultaneously, the results would give a more accurate picture. This multivariate approach is represented in Figure 2.4 and is considered below in detail.

2.2 Joint analysis of features

A joint analysis of embryo, oocyte and follicle features typically involves the grouping of embryos based on an observed development outcome (such as whether they resulted in a pregnancy or not on transfer). The features defining the embryos, oocyte and follicle of origin belonging to each group are then obtained from laboratory records. In order to identify the relationship between all these features and developmental potential, the data are analysed to determine the combination of features occurring most frequently in the pregnant group that can differentiate it from the non-

pregnant group. The features that are important and their relationship to developmental potential can be obtained by standard methods of multivariate analysis, such as linear regression or discriminant analysis.

Multivariate approach

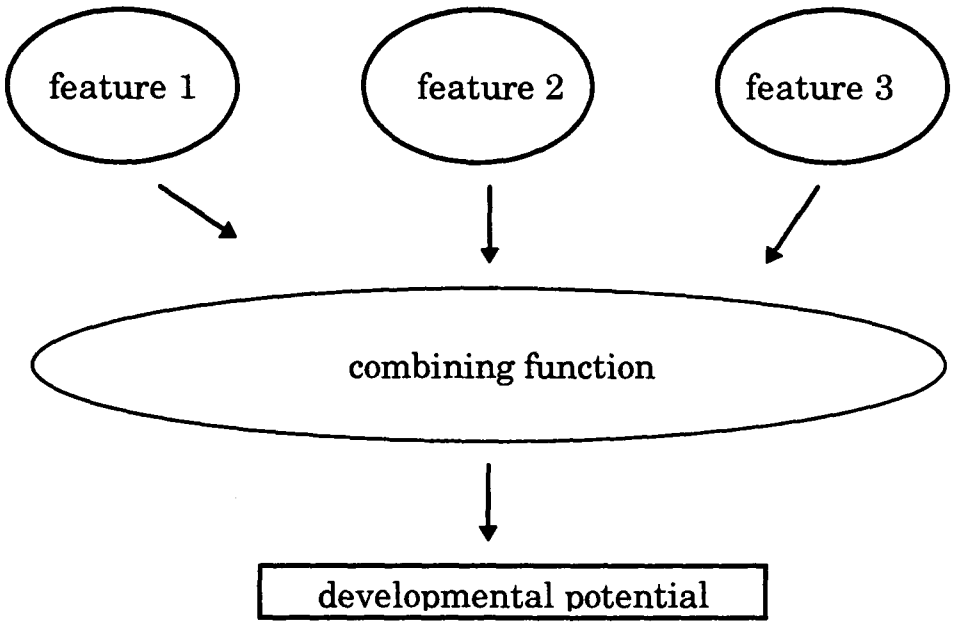


Figure 2.3 Multivariate approach: features are selected and combined and their joint influence investigated

To the author's knowledge, only two studies using this approach have been reported so far. Nayudu *et al* (1987) considered 3 groups of embryos based on the outcome of transfer. These were single embryo transfers resulting in a) normal pregnancy (i.e. term or near term infant); b) abnormal pregnancy (i.e. biochemical, blighted ovum or first trimester abortion); and c) no pregnancy. The features used in the study related both to each embryo as well as the follicle of its origin and are detailed in Table 2.4.

Table 2.4 Features analysed in the 'multivariate' study by Nayudu *et al* (1987)

No.	Feature included	Details
1	Follicular fluid components	complement C3, IgG2, total protein and Proteoglycans
2	Embryo appearance rating	based on regularity and distinct appearance of the blastomeres, fragmentation, appearance of the cytoplasm and any abnormalities
3	Follicular fluid volume	

The features for the three groups of embryos were analysed using the statistical technique of discriminant analysis (for a clear explanation, see SPSS Advanced Statistics User's Guide, 1990). The relationship between the features and the development outcome was expressed as a mathematical equation. It was possible to use this equation prospectively to identify whether an embryo has the potential to result in a normal pregnancy, abnormal pregnancy or no pregnancy.

The method described above gives a simple 'yes' or 'no' answer to whether an embryo transferred would result in a pregnancy. However, what is the probability that a 'yes' recommendation would really result in a pregnancy or how accurate are the identified relationships? Discriminant analysis is not capable of answering such a question directly but the technique of Logistic Regression can. In a subsequent study, Nayudu *et al* (1989) used this technique to investigate the role of characteristics of the follicle, sperm and patient (e.g. hormone levels, etc. in relation to that treatment cycle) in predicting the transfer outcome.

The authors found that three features were important in predicting the probability that a set of embryos transferred to a patient would result in a normal pregnancy - the mean level of Complement C3 in the follicular fluid, the first day of the cycle showing a sustained estradiol rise (E_2FD) and the number of transferred embryos that originated from follicles satisfying an idealised 'pregnancy follicle' criterion (that is, those follicles that had an $alpha_1$ anti-trypsin ratio > 0.8, low proteoglycans and follicular fluid volume between 2.5ml and 6.5 ml). Again, the relationship between these features and the developmental potential was expressed as a mathematical equation that could be used to calculate the probability that a given batch of embryos when transferred to a patient would result in a normal pregnancy.

Both discriminant analysis and logistic regression are flexible statistical techniques, making it easy to revise the equation should additional features become available. They can handle a large number of features from different sources. They do however, have three major shortcomings. Firstly, the relationships are expressed as mathematical equations rendering biological interpretation difficult. This considerably hinders their practical application. Secondly, these techniques make statistical assumptions regarding the data. For example, linear regression assumes that the features being analysed have a linear relationship to the developmental outcome. In IVF data however, such linear relationships may not hold true. For example, it is known that as follicle size increases, oocyte maturity and quality improve, but beyond a critical size, due to

post-maturity, the quality of the contained oocyte and therefore developmental potential of associated embryos is poor. Finally, the methods do not account for interaction between features. As an example from a different field, consider the interaction between gender and age when considering their relationship with height. When young, females grow more rapidly than males. At later ages, males grow faster. To understand correctly the relationship between age and height, the interaction between gender and age must be taken into account. If such interactions exist amongst the embryo, oocyte and follicle features, techniques like regression and discriminant analysis could give misleading results.

Analysis techniques used thus far have therefore not been ideal. Methods are required which overcome the shortcomings listed earlier, whilst retaining the ability to perform a multivariate analysis. Precisely such concerns have seen the development of 'Decision Tree' analysis techniques in the area of Machine Learning in Computer Science. As with discriminant analysis and logistic regression, these methods can analyse a large number of features simultaneously and have the flexibility to include new features easily. In addition, they take into account conditional dependencies between features. Figure 2.4 represents the ability of decision trees to jointly analyse features, as well as consider interactions and contrasts it against regression. Finally, and importantly, results are expressed in the form of easily understandable rules (Michie, 1991).

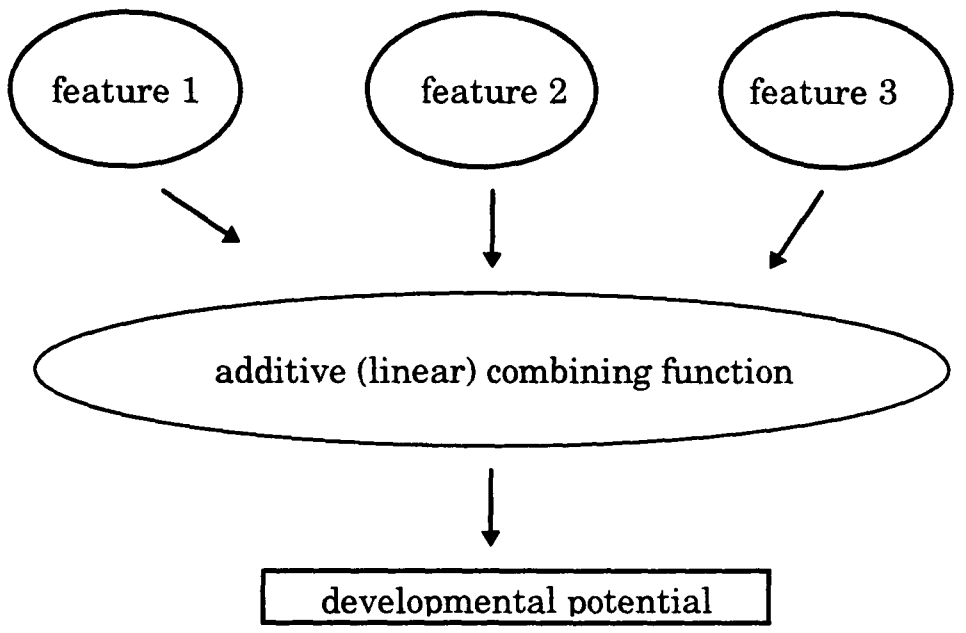


Figure 2.4 Logistic regression versus decision tree approach

Decision tree techniques have been used successfully in diverse areas such as predicting the likelihood of severe thunderstorms (Zubrick, 1988), guiding the landing of the Space Shuttle (Michie, 1984), evaluating the credit-worthiness of clients applying for a credit card (Michie, 1988), ECG interpretation (Bratko *et al*, 1989) and location of primary tumours (Clark

and Niblett, 1989). Success in biological and other different areas, along with the desirable properties mentioned earlier provides the impetus to selecting decision-trees as the preferred method of analysis in this dissertation.

2.3 Summary

This review has looked at various methods of investigating the relationship between embryo-related features and developmental potential. Relatively simple methods that exclusively analysed the individual embryo or oocyte or follicle features were first examined. However, since development is naturally influenced by a range of features, the ability to investigate multiple features taking into account their joint influence on each other is very important. Studies that have used methods capable of combining information from a large number of features were then described. The statistical methods used in these studies can be automated easily, and are flexible enough to incorporate new features as and when they become available. In order to derive the maximum information from such investigations however, it is essential for conditional dependencies and interactions between features to be accounted for as well as for the results to be expressed in an easily understandable form, rather than as mathematical equations. This dissertation explores the use of 'decision tree techniques' which possess such properties, towards a thorough investigation of

relationships between features and developmental potential of human IVF embryos.

Chapter Three. Background And Methodology

3.1 INTRODUCTION 38

3.2 ORGANISATION OF THIS CHAPTER 38

3.3 OXFORD IVF UNIT: MATERIALS 39

3.3.1 Ivf Drug Regime 39

3.3.2 Media 41

3.3.3 Microscopes 42

3.3.4 Plastic And Glassware 43

3.4 OXFORD IVF UNIT: METHODS 45

3.4.1 Oocyte Retrieval 45

3.4.2 Insemination (Day 0) 49

3.4.3 Monitoring Fertilisation (Day 1) 50

3.4.4 Monitoring Early Embryo Development (Day 2) 53

3.4.5 Criteria Used To Select Embryos For Transfer 56

3.4.6 Embryo Transfer 60

3.4.7 Monitoring Embryo Development In Vitro 60

3.5 DECISION TREES AND RULES 61

3.5.1 Use Of C4.5 For Decision Tree Analysis 63

3.5.2 Data Used To Construct Trees And Rules 65

3.5.3 Extracting Feature Combinations To Build Trees And Rules 66

3.5.4 What Is The Accuracy Of The Rules? 69

3.5.5 Simplification Of Trees And Rules 75

3.6 CONSTRUCTING TREES AND RULES IN THIS DISSERTATION 77

3.7 SUMMARY 79

3. BACKGROUND AND METHODOLOGY

3.1 Introduction

In Chapter 2, the concept of using the statistical technique of decision trees to investigate the relationships between embryo-related features and the developmental potential of human IVF embryos was introduced. This chapter details the materials and methods used in IVF treatment and embryo culture and describes the decision tree technique.

3.2 Organisation of this chapter

The first half of this chapter begins with a discussion on materials used in IVF (Section 3.3). This is followed by a detailed description of methods involved in treatment in Section 3.4. The assessment of follicles, oocytes and embryos is described in order to provide a comprehensive understanding of the features used in analyses in later chapters. A brief description of the procedure for culturing spare embryos *in vitro* completes the first part. This is of relevance to material presented in Chapter 6.

The second half of this chapter provides a lay introduction to decision tree analysis in Section 3.5 followed by a description of the methods used when applying the decision tree technique to the analysis of IVF data (Section 3.6).

3.3 Oxford IVF Unit: Materials[1]

Most practices described in this chapter are in current use in the Oxford IVF Unit. Although minor methodological changes have taken place since the collection of data for this thesis, reference to the materials and methods in the past tense does not reflect any major alterations to IVF practice.

3.3.1 IVF drug regime

Ovarian stimulation

Normally ovulation results in the release of a single oocyte during the menstrual cycle. Patients undergoing IVF treatment however, are administered drugs to induce superovulation. The drug regime used here involved two steps, namely, pituitary down regulation and follicular stimulation. Pituitary down regulation was achieved by the administration of the gonadotrophin releasing hormone (GnRH) analogues, Buserelin (Suprefact; Hoechst, Hounslow, UK Ltd.) or Nafarelin (Synarel; Syntex Pharmaceuticals Limited, England). Both were administered by the intra-nasal route, Buserelin at a dose of 200µg five times daily or Nafarelin at a dose of 400µg twice daily. GnRH analogue administration commenced on the 21st day of the menstrual

---

[1] The materials section has been written with close reference to pages 23-25, in Section 2.2 Materials and Methods, of the D.Phil. thesis, 'Human trophectoderm biopsy and its application', submitted by A.Dokras to the University of Oxford, 1992

cycle and down regulation occurred a minimum of 18 days (and usually three weeks) after the beginning of administration of the drugs. The criteria for down-regulation were serum oestradiol levels < 20pg/ml and the absence of any ovarian follicle > 9mm in diameter as assessed by ultrasound. Administration of the GnRH analogue, however continued until the patient received an injection of human chorionic gonadotrophin (hCG) usually 4-5 weeks after the onset of drug treatment.

Once pituitary down regulation was achieved, follicular stimulation was obtained by the intramuscular injection of 150-300 IU of the human menopausal gonadotrophin, Pergonal (75 IU follicle stimulating hormone + 75 IU human luteinising hormone; Serono Laboratory, Welwyn Garden City, UK) daily for 9-13 days[2]. The response of the patient was monitored from the 8th day of follicular stimulation onwards by the measurement of blood oestradiol and progesterone levels together with the ultrasound appearance of the follicles. When the criteria for adequate follicular development were achieved (at least 3 follicles with a mean diameter ≥ 16mm or 5 follicles ≥ 15mm), 10,000 IU of hCG (Profasi, Serono Laboratory) was injected intramuscularly to induce the final stages of maturation of the oocytes. Oocyte recovery was scheduled for 35 hours following the administration of hCG.

---

[2] Metrodin (75 IU pure Follicle Stimulating Hormone; Serono Laboratories UK Limited) in a daily dose of 225IU was administered to women with high luteinising hormone levels due to the presence of polycystic ovaries as well as to women who showed over-stimulation with very high levels of oestradiol when treated with Pergonal in a previous treatment cycle.

Analgesia

Retrieval was performed under sedation and analgesia induced by the intravenous administration of 50mg Pethidine (Roche Products Limited, Welwyn Garden City, England) and 5mg Diazemuls [Diazepam (Dumex Limited, Princes Risborough, Bucks, UK.)].[3]

3.3.2 Media

The stock solution of the medium used for flushing oocytes from follicles as well as for culturing oocytes and embryos was prepared every fortnight, filter sterilised and stored at 4°C. Ultra pure pyrogen free water (Milli-Q UF Plus Water Purification System, Millipore (UK) Ltd, Watford, UK) was utilised for preparation.

Flushing medium

Heparinised Hepes buffered Tyrode's 6 (T6) 'flushing medium' was used to flush oocytes from follicles during oocyte retrieval (see Table 3.1).

Culture medium

T6 supplemented with 10% heat inactivated autologous maternal serum and equilibrated overnight at 37°C in a Heraeus B 5060 CO_2 incubator (Heraeus Equipment Ltd, Essex, UK) set at a humidified atmosphere of 5% CO_2 in air was used to culture the oocytes and embryos. The same

---

[3] Hypnovel (Midazolam; Roche Products Limited, Welwyn Garden City, England) was used instead of diazepam in patients with a low pain threshold.

medium was used to load embryos into the catheter during embryo transfer as well as to culture spare embryos *in vitro* until they were 14 days.

Serum

Serum supplement for the culture of each woman's embryos was obtained from the peripheral venous blood collected on the 8th day of her ovarian stimulation regime. Serum was obtained by centrifuging the blood at 1000g for 10 min in 10ml sterile polystyrene centrifugation tubes (Falcon 2001, New Jersey, USA). Erythrocytes and leucocytes were thus precipitated and the plasma supernatant allowed to clot to give the serum. The serum was then heat inactivated at 56°C for 30min and filter sterilised (0.22μm filter, Millipore (UK) Ltd., Watford, UK).

3.3.3 Microscopes

Early embryo development was visualised through a stereo dissecting microscope (M8, Wild, Heerbrugg, Switzerland) which was placed in a laminar flow hood (Gelaire - HF 36, Flow Laboratories Scotland). The microscope stage was fitted with an electric filament (Cryogenic Technology, Inc, USA) adjusted to maintain the temperature of culture medium at 37° C. A Wilovert inverted phase contrast microscope (Leica, Milton Keynes, UK) fitted with a heated stage was used to visualise the embryos from the blastocyst stage onwards.

Table 3.1 Composition of Tyrode's 6 'flushing' and 'culture' media

Stock A = solution 1 + solution 2

Solution 1

Ingredients	Weight (grams)
NaCl	5.991
Na Lactate	4.872
D-Glucose	1.048
KCl	0.111
$MgCL_2.6H_2O$	0.101
Na_2HPO_4	0.054
Na Pyruvate	0.054
Penicillin G	0.063
Phenol Red	0.005
Dissolve in 500ml water	

Solution 2

Ingredients	Weight (grams)
$CaCl_2.2H_2O$	0.274
Dissolve in 100 ml water	

Stock B

Ingredients	Weight (grams)
NaHCO3	2.101
Dissolve in 100ml water.	

T6 1 litre = 900ml Stock A + 100ml Stock B (Osmolality 285mOsmol/kg)

T6 'flushing medium'

4920ml Stock A
79ml Stock B
100ml 1M Hepes (pH 7.2-7.4)
30 vials of heparin (5000units/0.2ml)

T6 'culture medium'

T6
10% heat inactivated, filter sterilised autologous maternal serum

3.3.4 Plastic and glassware

All plasticware used for IVF treatment and embryo culture was sterile. Fluid was aspirated from the follicles into 10ml sterile centrifugation

tubes (Falcon 2001). Prior to transfer, oocytes and embryos were incubated in 2 ml medium in central wells of organ tissue culture dishes (60 × 15 mm, Falcon 3037). Following the transfer of 3 selected embryos, the spare embryos assigned for research were transferred to 1ml medium contained in individual wells of 24 well Nunclon plates (Nunc, Delta, Denmark).

The glassware used for embryo handling was tissue culture washed. This involved washing the glassware twice for 10min in an ultrasonic bath (Decon, Ultrasonics LTD, England) containing a 0.25% solution of 7X detergent in water that had been deionised by reverse osmosis (Milli-RO 10TS, Reverse Osmosis System, Millipore (UK) Ltd, Watford, UK). In order to ensure proper removal of the detergent, the glassware was then washed six times with Milli Q UF Plus water (see 3.3.2) and hot air sterilised at 240°C overnight.

Glass Pasteur pipettes sterilised by this method and then resterilised overnight after transfer into aluminium boxes were used to handle embryos. To obtain a tip of the required shape, the part of the pipette close to the tip was softened by rotating over the flame of a gas burner. As the glass melted the two ends were pulled apart sharply so as to cause the pipette to break at this point, and withdrawn from the flame. The edge of the pipette was then squared by snapping the broken glass tubing against it. Before use, the tip was checked under the dissecting microscope to ensure that it was smooth.

3.4 Oxford IVF Unit: Methods

3.4.1 Oocyte retrieval

Oocyte retrieval (ovum pick-up) was performed by vaginal ultrasound guided oocyte aspiration 35 hours after the administration of hCG. A premature pick-up would result in the retrieval of inadequately matured oocytes. In the case of a delay, spontaneous ovulation with follicular rupture and oocyte release could occur, making retrieval difficult.

Once the vaginal ultrasound probe was in position, the follicles appeared on the ultrasound screen as dark echogenic circular areas separated by white areas of ovarian tissue. Before aspiration, the diameter of the follicle to be aspirated was measured on the ultrasound screen. Usually the diameter was measured at two different points and the average diameter in mm recorded on the oocyte recovery record sheet. Occasionally, to save time as well as to reduce the patient's discomfort, only a single diameter was measured or in the case of small follicles, a rough estimation of the follicle size was made based on the coarse scale available on the ultrasound screen. After recording its size, the follicle was aspirated and the follicular fluid (ff) drained out into a warm sterile test tube. Aspiration was continued until the follicle was seen to collapse completely on the ultrasound screen and no more fluid was obtained. The appearance and the volume of the ff collected in the test tube was recorded. The fluid was then transferred into an organ tissue culture dish and scanned with a microscope for the presence of an oocyte.

Once an oocyte was found in the ff its quality was assessed based mainly on features relating to the cumulus-corona complex (features 4-7, Table 3.2 and Plate B) and it was classified as belonging to one of 5 grades (details of oocyte grading as in Table 3.2, feature 8). The oocyte was then transferred to another equilibrated organ tissue culture dish containing 2ml of T6 supplemented with 10% autologous maternal serum, and incubated for 4-7 hours prior to insemination.

Table 3.2 gives details of the features that were recorded in the oocyte recovery sheet of each patient, during and following the recovery of an oocyte from a follicle. At this stage the oocyte (secondary oocyte) is normally arrested in the metaphase of the second meiotic division and the first polar body may be visualised. The second meiotic division with the formation of the ootid and the second polar body is only completed when the oocyte is activated by the entry of the sperm into the ooplasm after insemination.

Table 3.2 Features describing oocyte and associated follicle

No.	Feature	Range of values normally observed	Embryologists' interpretation of these values
1	follicle size	10-38mm	15-20mm = mature oocyte ≤14mm = immature oocyte >20mm = post-mature oocyte in most cases.
2	ff volume	0.5-9.0ml	as the volume increased, it indicated larger follicle size and therefore a more mature oocyte. A very high volume indicated post maturity.

3	ff appearance	clear yellow very slightly bloody slightly bloody bloody very bloody	ff appearance did not reflect oocyte quality. However a clear yellow appearance indicated that the needle was definitely aspirating a follicle. Presence of blood indicated trauma to the surrounding ovarian tissue. A very bloody appearance warned of the possibility of entry into a blood vessel. A chocolate coloured appearance indicated entry into an endometriotic cyst.
4	oocyte cumulus description	dark (immature)[4] slightly dark dark (just post mature)[4] very slightly postmature slightly postmature post mature	the compact cumulus in an immature oocyte resulted in a dark appearance. the compactness and therefore the darkness decreased in a mature oocyte. the degeneration of cumulus in a postmature oocyte also resulted in a dark appearance but this "darkness" differed in appearance from the "darkness" associated with an immature oocyte. increase in post maturity was associated with blood streaking.
5	oocyte cumulus amount	very little medium lots	normally the amount of cumulus was found to increase with the degree of maturity. Oocytes associated with very little cumulus were considered immature and those associated with lots of cumulus, postmature.

[4]The IVF records use the same keyword "dark" in 2 different contexts. It is impossible to distinguish from the records, between the darkness related to immaturity and that related to postmaturity.'

6	oocyte corona description	okay slightly dark dark border very dark	an 'okay' corona was one that was well spread out and therefore appeared light. This signified a mature oocyte. Increasing darkness i.e. slightly dark, dark and very dark indicated an increase in the compactness of the corona and an increasing degree of immaturity. Border indicated that the appearance was quite dark (between dark and very dark) but it was not clear if this was due to the compactness of an immature corona or it was a spread out mature corona associated with a darkish appearance
7	oocyte corona maturity	very immature immature mature	a very compact dense corona indicated a very immature oocyte while the corona of an immature oocyte was fairly compact. A fairly spread out corona indicated a mature oocyte.
8	egg grade	grade 1 (post mature) grade 2 (mature) grade 3 (borderline immature) grade 4 (immature)	egg grade indicated the maturity status of the oocyte and was given by the embryologists to the oocyte following an assessment of the features describing the cumulus-corona complex (features 4-7).
9	darkness of the ooplasm	slightly dark dark very dark	any degree of darkness indicated an abnormal appearance and therefore poor oocyte quality.
10	granularity of the ooplasm	slightly granular granular very granular	any degree of granularity indicated an abnormal appearance and therefore poor oocyte quality.
11	elongation of the ooplasm[5]	slightly elongated elongated very elongated	any degree of elongation indicated an abnormal appearance and therefore poor oocyte quality.

[5] Sometimes though both the zona and the oocyte were elongated, this was not recorded. Either the zona or the oocyte were recorded as having an elongated appearance.

12	shrunken appearance of the ooplasm	slightly shrunken shrunken very shrunken	a shrunken appearance indicated an abnormal appearance and therefore poor oocyte quality.
13	vesicles in the ooplasm	1,2,3...vesicles	presence of vesicles (light circular areas) in the ooplasm indicated poor oocyte quality.
14	refractile bodies in the ooplasm	1,2,3...refractile bodies	presence of refractile bodies (dark circular areas) indicated poor oocyte quality. Such oocytes rarely fertilised.
15	misshapen ooplasm	slightly misshapen misshapen very misshapen	a misshapen appearance indicated poor oocyte quality.
16	thickness of the zona pellucida	thick zona very thick zona	a thick zona was considered a hindrance to fertilisation and hatching and therefore an indicator of poor oocyte quality.
17	elongation of the zona pellucida[6]	slightly elongated zona elongated zona very elongated zona	an elongated zona pellucida indicated poor oocyte quality.

3.4.2 Insemination (day 0)

Standard semen analysis including an assessment of volume, pH, sperm concentration, motility, progressive motility, morphology, migration, evidence of infection and agglutination was performed on the semen sample which was collected 2½ to 3h before insemination. The sample was prepared for insemination by the swim-up technique from a washed

[6] Same as footnote 5

pellet (Mortimer, 1985). It was then added at a concentration of 1×10^5 - 2×10^5 spermatozoa/ml to the culture dish containing a single oocyte (day 0). A low total sperm count due to low semen volume or low sperm density however, necessitated the insemination of multiple oocytes in a single dish. In such cases an attempt was made to pool oocytes originating from follicles of a similar size, and possibly therefore a similar maturity status, in the single dish.

3.4.3 Monitoring fertilisation (day 1)

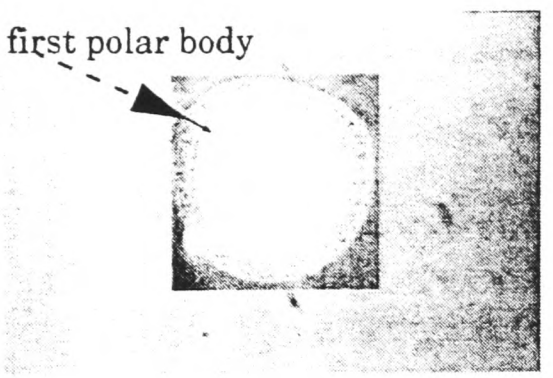
Oocytes were examined for the occurrence of fertilisation 15.5 to 18 hours after insemination. Fertilisation was indicated by the visualisation of pronuclei and, if possible, two polar bodies (see Plate B). To allow a clearer view, the cumulus corona complex was stripped by pipetting the oocyte repeatedly through a fine-tipped, flame-polished pipette and then transferring the oocyte to a new tissue culture dish containing equilibrated T6 + 10% autologous maternal serum. A number of features were recorded at the time of fertilisation check and Table 3.3 gives details of these features and their significance. Following transfer of the oocyte/embryo to the new dish, the insemination dishes were visualised under a phase contrast microscope in order to record the cumulus dish attachment (detailed in Table 3.3, feature 11) and overnight sperm survival (Table 3.3, feature 12). If no evidence of fertilisation was observed, the unfertilised oocytes were observed periodically over the next 24h for detection of delayed fertilisation.

Table 3.3 Features recorded when checking oocytes for fertilisation (day 1)

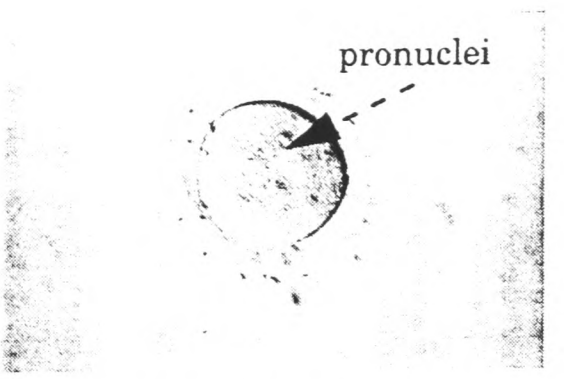
No	Feature	Range of values	Embryologists' interpretation of values
1	fertilisation status	single cell + 1 polar body (pb)	indicated that no fertilisation had taken place.
		1 pro-nucleus + 2 pb	indicated abnormal fertilisation.
		1 pro-nucleus + 1 pb	indicated parthenogenic activation.
		2 pronuclei (+ possibly 2pb)	indicated fertilisation of the oocyte by a single sperm.
		3,4... pronuclei	indicated fertilisation of the oocyte by multiple sperm.
		syngamy	although neither pronuclei nor evidence of cleavage was present, visibility of the 2 polar bodies indicated that fertilisation had taken place.
2	darkness of the cytoplasm	slightly dark dark very dark	any degree of darkness indicated an abnormal appearance and therefore poor embryo quality.
3	granularity of the cytoplasm	slightly granular granular very granular	any degree of granularity indicated an abnormal appearance and therefore poor embryo quality.
4	elongation of the embryo[7]	slightly elongated elongated very elongated	any degree of elongation indicated an abnormal appearance and therefore poor embryo quality.

[7]Sometimes though both the zona and the embryo are elongated, this may not be recorded. Either the zona or the embryo may be recorded as having an elongated appearance.

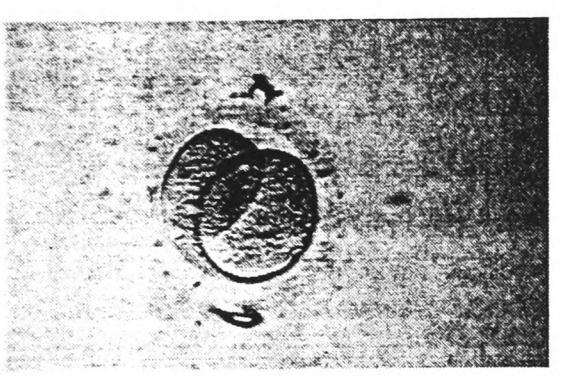
5	shrunken appearance of the cytoplasm	slightly shrunken shrunken very shrunken	a shrunken appearance indicated an abnormal appearance and therefore poor embryo quality.
6	vesicles in the cytoplasm	1,2,3... vesicles	presence of vesicles in the cytoplasm indicated poor embryo quality.
7	refractile bodies in the cytoplasm	1,2,3... refractile bodies	presence of refractile bodies was very rare at this stage as oocytes with refractile bodies very rarely fertilised.
8	misshapen cytoplasm	slightly misshapen misshapen very misshapen	a misshapen appearance indicated poor embryo quality.
9	thickness of the zona pellucida	thick zona very thick zona	a thick zona was considered a hindrance to fertilisation and hatching and therefore an indicator of poor embryo quality.
10	elongation of the zona pellucida[7]	slightly elongated zona elongated zona very elongated zona	an elongated zona pellucida indicated poor embryo quality.
11	cumulus dish attachment	massive lots moderate little none	cumulus dish attachment refers to the amount of cumulus found attached to the dish. Stickier cumulus, resulting in the adherence of larger amounts to the dish, indicated oocyte maturity.
12	sperm survival	excellent good okay many dead poor all dead	sperm survival indicates the overnight sperm survival and was assessed based on sperm motility. A low motility indicated the possible presence of antisperm antibodies in the maternal serum used, favouring the use of a human serum albumin supplemented medium in a subsequent cycle



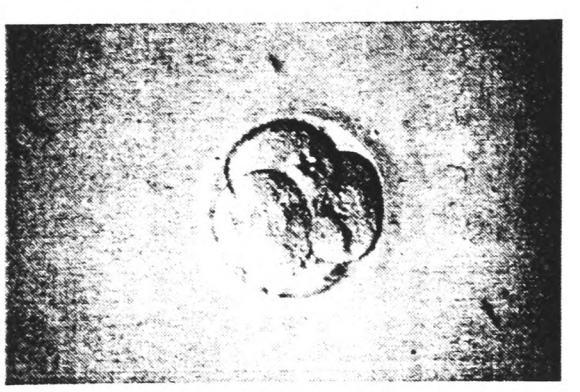
unfertilised
oocyte (day 0)



fertilised oocyte
(day 1)



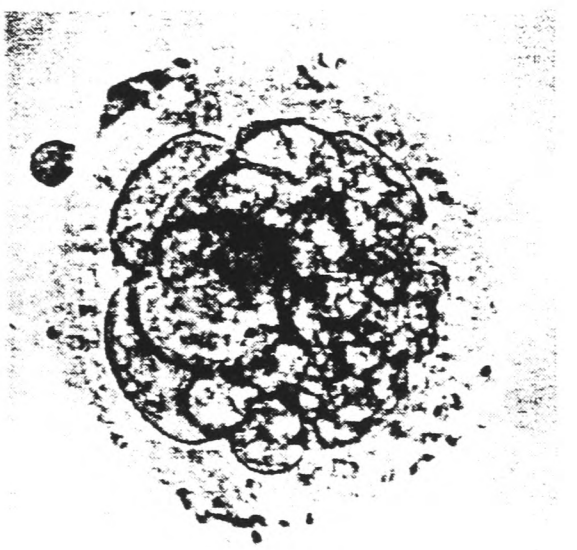
cleaving embryo - 2 cell
stage (day 2)



cleaving embryo - 4 cell
stage (day 2)



embryo with elongated
zona, fragmentation
and irregular cells



embryo with 'lots of
fragments' and 'very
irregular' cells

Plate B. Stages of development up to day 2 and some characteristics observed on check 1 and check 2

3.4.4 Monitoring early embryo development (day 2)

Embryos were checked twice on the day of transfer. Transfer typically took place on the second day following insemination (day 2). The first check was carried out in the morning, usually between 8.00 and 10.00 a.m. (check 1, day 2). The second check was performed just before embryo transfer, usually in the afternoon (check 2, day 2). On weekends however, embryo transfers tended to take place in the mornings. Therefore a single check before transfer was carried out.

Table 3.4 gives details of the features recorded following embryo checks in the embryo transfer sheet and their significance as interpreted by the embryologists. Plate B shows the stage of development normally achieved by the embryo by day 2.

Embryo grading

Embryo grade is one of the features mentioned in Table 3.4 (feature 14) and was assessed as follows. At the time of the embryo check performed just before embryo transfer (check 2, day 2), embryos were classified into 5 grades based on the cell number, fragmentation, regularity, darkness and granularity of the cytoplasm. Although guidelines were available, the grading was to a certain extent subjective. Details of the grading are shown in Table 3.5.

Table 3.4 Features recorded following embryo checks

No.	Feature	Range of values	Embryologist's interpretation of these values
1	cell (or blastomere) number	single cell with 1 pb	indicated that no fertilisation had taken place
		syngamy	neither pronuclei nor evidence of cleavage was present, but 2 polar bodies were visible indicating that fertilisation had taken place
		2,3,4... cells	cleavage had taken place. 2 polar bodies may or may not be visible. If a single pb was visible and this embryo had originated from an oocyte that earlier showed evidence of normal fertilisation, it could just indicate that denaturation of one of the polar bodies had taken place. Alternatively if this embryo had originated from an oocyte which at the time of the fertilisation check showed no evidence of fertilisation, it could indicate a) the oocyte had been parthenogenically activated or b) it could have undergone normal fertilisation which was delayed with the subsequent denaturation of one of the polar bodies.
2	Regularity of cells in relation to one another (see Plate B)	regular	an embryo with regular cells was considered an embryo of good quality. As the degree of irregularity increased, embryo quality was considered to progressively worsen.
		slightly irregular	
		irregular	
		very irregular	
3	fragmentation in the perivitelline space (see Plate B)	no fragments	embryos with no fragments were considered embryos of good quality. As the amount of fragmentation increased quality was considered to progressively worsen.
		few tiny fragments	
		few small fragments	
		few fragments	
		fragments	
		lots of fragments	
4	visibility of nuclei inside cells	nuclei visible	embryos with visible nuclei were preferred for embryo transfer

5	darkness of the cytoplasm of the cells	slightly dark dark very dark	any degree of darkness indicated an abnormal appearance and therefore poor embryo quality.
6	granularity of the cytoplasm of the cells	slightly granular granular very granular	any degree of granularity indicated an abnormal appearance and therefore poor embryo quality.
7	elongation of the embryo[8]	slightly elongated elongated very elongated	any degree of elongation indicated an abnormal appearance and therefore poor embryo quality.
8	shrunken appearance of the cells	slightly shrunken shrunken very shrunken	a shrunken appearance indicated an abnormal appearance and therefore poor embryo quality.
9	vesicles in the cytoplasm of the cells	1,2,3... vesicles	presence of vesicles in the cytoplasm indicated poor embryo quality.
10	refractile bodies in the cytoplasm of the cells	1,2,3... refractile bodies	this was an extremely rare occurrence as it was observed that oocytes with refractile bodies very rarely fertilised and even more rarely did they go on to cleave.
11	misshapen ooplasm	slightly misshapen misshapen very misshapen	a misshapen appearance indicated poor embryo quality.
12	thickness of the zona pellucida	thick zona very thick zona	a thick zona was considered a hindrance to fertilisation and hatching and therefore an indicator of poor embryo quality.
13	elongation of the zona pellucida[8]	slightly elongated zona elongated zona very elongated zona	an elongated zona pellucida indicated poor embryo quality.

[8]Sometimes though both the zona and the embryo are elongated, this may not be recorded. Either the zona or the embryo may be recorded as having an elongated appearance.

14	embryo grade	grade 1 grade 2 grade 3 grade 4 grade 5	embryos were classified into 5 grades at the time of the second check, just before transfer on day 2. Details of the grading procedure are given below in Table 3.5

Details of the criteria for allotting an embryo to a particular grade are outlined in Table 3.5 below.

Table 3.5 Embryo grading (Grade 1 indicates best grade)

Embryo grade	Description
grade 1	regular 4 - 5 cell ± few fragments or slightly irregular 4 - 5 cell ± few fragments
grade 2	regular ≥ 6 cell ± few fragments or slightly irregular ± few fragments or few tiny fragments
grade 3	irregular ≥ 6 cell ± few fragments or regular ≥ 6 cell + lots of fragments or dark or granular ooplasm or irregular 4 - 5 ± fragments or few fragments or regular 4 - 5 + lots of fragments or regular 2 - 3 ± few fragments or slightly irregular 2 - 3 ± fragments
grade 4	irregular 2 - 3 ± fragments or few fragments or regular 2 - 3 ± lots of fragments
grade 5	single cell or ball of fragments

3.4.5 Criteria used to select embryos for transfer

In order to minimise the risk of multiple pregnancy, a maximum of 3 embryos were chosen for transfer (HFEA, Code of Practice, 1991 onwards)

A number of criteria were taken into consideration by the embryologists when selecting embryos for transfer. The selection method was developed and refined based on past experience of transferring embryos with certain characteristics, as well as biological insights obtained from a number of studies investigating relationships between particular features and the developmental potential of embryos. The selection procedure (i.e. the features) used however was not corroborated by an independent study.

In order of importance, the criteria taken into consideration while selecting these embryos for transfer of those available, were:

1. **Number of pronuclei.** Ordinarily, only bipronucleate embryos were considered for transfer. If, in a given batch of embryos, the quality of the bipronucleate embryos was assessed as not being as good as that of any available unipronucleate or parthenogenic embryos, the bipronucleates would still be preferred. Unipronucleate or parthenogenic embryos were considered only in situations where 3 bipronucleate embryos were not available. Multipronucleate embryos were never considered for transfer.

2. **Embryo grade.** Grade 1 embryos were most preferred followed by grade 2 and 3. Embryos belonging to grades 4 and 5 were considered for transfer only if embryos belonging to grades 1-3 were unavailable.

3. **Follicle size.** Embryos resulting from oocytes that were recovered from follicles between the sizes of 16 - 18 mm were preferred. In fact a grade 3 embryo that was associated with a follicle of size between 16 -

18 mm could be preferred over an embryo of grade 1 that had an associated follicular size outside the preferred range.

4. **Darkness of corona.** Criteria 1-3 being the same for a group of embryos, those embryos that were associated with an 'okay' i.e. lighter corona were preferred. However, in practice very few embryos were associated with a light corona and a 'slightly dark' corona was acceptable. Given a choice, embryos originating from oocytes with a 'dark', 'border line' or 'very dark' corona were not chosen.

5. **Visibility of nuclei.** Criteria 1-4 being the same, embryos that showed visible nuclei in the cells were preferred.

6. **Abnormalities of the ooplasm, cytoplasm or zona pellucida.** Embryos that showed abnormalities associated with the ooplasm or cytoplasm i.e. darkness, granularity, elongation, odd shapes, shrunken appearance, presence of vesicles or refractile bodies were usually not chosen. Further, even if an embryo had a very good grade and ideal appearance of the corona and an ideal follicle size, the association of abnormalities like elongation or thickening of the zona pellucida reduced the chance of the embryo being selected for transfer.

7. ***In vitro* developmental potential of spare embryos from a previous cycle.** In case of patients who had previously had a treatment cycle at the IVF unit, the records of the *in vitro* development of spare embryos in those cycles were referred to. A note was made of the grade, follicle size and appearance of the corona associated with

embryos that proceeded to form blastocysts. Embryos that showed properties similar to these, were preferred over others while selecting embryos for transfer.

While some of these criteria could be assessed simultaneously rather than progressively, the selection procedure can be represented by the following flow-chart (Figure 3.1). Note that factors relating to the earlier cycle that have not been included in the diagram, were also taken into account for patients who had earlier received treatment in the clinic.

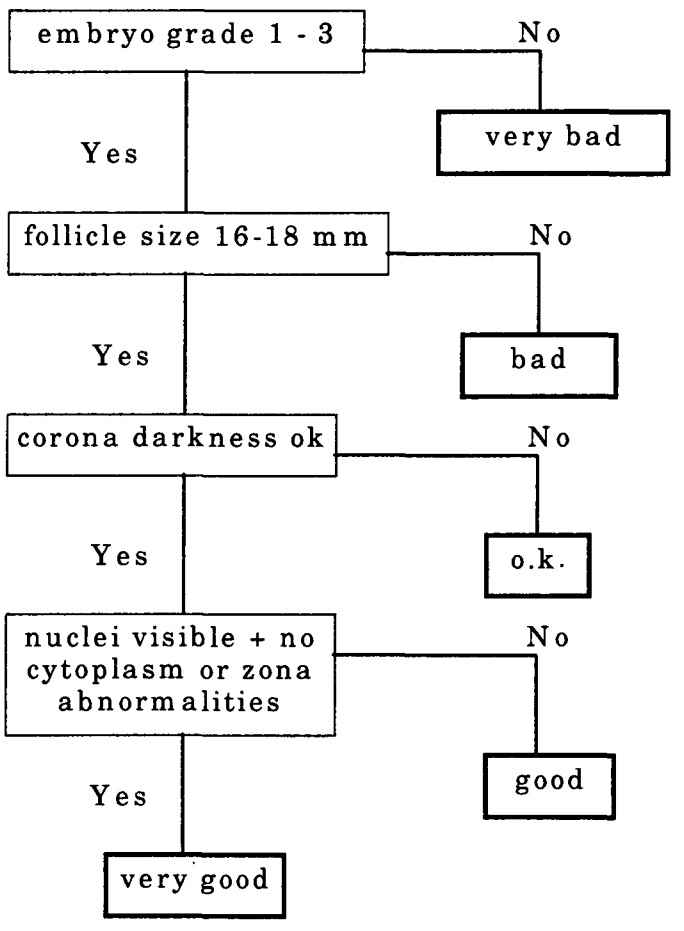


Figure 3.1 Method of selection of the 'best' embryos for transfer

The grading of oocytes and embryos and the selection of embryos was done by the two embryologists. Embryologists were trained rigorously and an attempt was made to ensure as much objectivity in the selection procedure as possible. However, since some of the features used were

subjective (e.g. amount of fragmentation, darkness of corona, embryo grades), the possibility of inter embryologist variation as well as variation in the selection by the same embryologist was unavoidable.

3.4.6 Embryo transfer

Approximately 48h after insemination, the batch of three selected embryos was loaded in 20-40μl medium in an intrauterine transfer catheter (Wallace, Essex, UK or Craft, Rocket of London, Watford, England) fitted with a 1ml syringe. The medium was positioned between two air bubbles to prevent drainage as well as to locate the position of the embryos. With the patient lying in a recumbent position, the cervix was exposed with a speculum and cleaned. The tip of the catheter was then introduced through the external and internal os of the cervix and the plunger of the syringe gradually depressed to expel the embryos into the uterine cavity. The catheter was gently withdrawn and the patient allowed to rest for half an hour before being discharged. The tip of the catheter was visualised under the microscope and flushed with medium after the transfer to ensure that the embryos were not retained in it.

3.4.7 Monitoring embryo development *in vitro*

Following transfer all bipronucleate spare embryos were donated for research only if the patient consented. These spare embryos were transferred to the IVF research laboratory and cultured in individual wells of 24 well Nunclon plates (Section 3.3.4). Every 24h the embryos

were assessed morphologically, (for details on microscope see 3.3.3) before being transferred to a new well containing 1ml of fresh equilibrated medium (see 3.3.2). The culture medium from the previous well was stored at minus 20°C and assessed later for levels of human chorionic gonadotrophin (data used in Chapters 3 and 6), immunosuppressive activity or pregnancy specific beta-1 glycoprotein secretion (detailed in Appendix B). If an embryo became attached to the culture dish, the medium was carefully removed and replaced with fresh equilibrated medium without dislodging the embryo.

Spare embryos obtained at the cleavage (4 - 6 cell) stage on day 2 progressed in culture through the 8 cell (day 3) and morula (day 4) stages gradually showing the appearance of a blastocoelic cavity and proceeding to the blastocyst stage (normally day 5 or 6). All embryos were cultured up to day 8, after which only blastocysts or vacuolated morulae (morulae showing vacuoles that could be mistaken for early cavitation but not progressing to blastocysts) remained in culture up to the time that the embryo was 14 days.

3.5 Decision trees and rules

The aim of this dissertation is to investigate the relationships between embryo, oocyte and associated follicular features and the developmental potential of human IVF embryos. Towards this end, the statistical technique of decision trees has been chosen to analyse data related to embryos. The advantages it has over traditional methods of analysis used

so far, have been outlined in Chapter 2. Briefly, the decision tree technique is capable of investigating a number of features taking into account any non-linear relationships and conditional interdependencies that may exist. The results are accurate as well as expressed in an easily understandable form.

The decision tree technique works by analysing data related to a sufficient number of specific examples and discovering the general combinations of features (i.e. patterns) characterising the examples in the group. The patterns are initially expressed as decision trees (hence the name of the technique) which can then be converted to an easily understandable form of statements or rules (see Section 3.5.3). This ability is exploited here by using the technique to analyse embryos with differing development potential. In this thesis, the examples are embryos and the data relates to features describing these embryos and the oocyte of origin and associated follicle. Depending on the outcome of transfer embryos are classified into different groups or classes (e.g. pregnant or not pregnant) of differing developmental potential. The tree technique is used to capture those combinations of embryo-related features that characterise each class.

Figure 3.2 gives a flow-chart of the steps to be followed in extracting the characterising patterns of features. The procedures are explained in greater detail in the sections sign-posted in the figure.

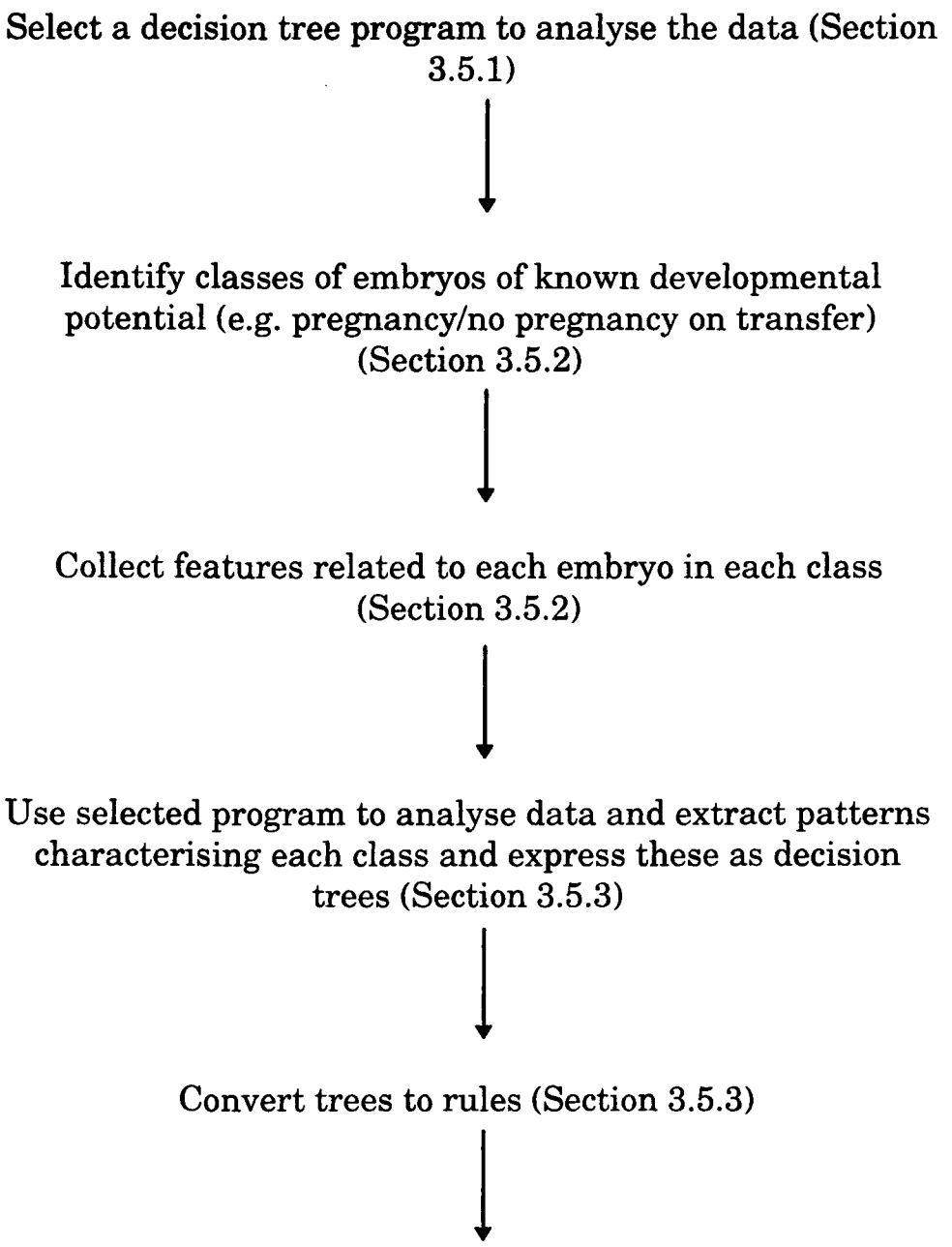


Figure 3.2 Procedure to obtain accurate patterns

3.5.1 Use of C4.5 for decision tree analysis

In this thesis, a suite of computer programs called C4.5 (Quinlan, 1993) was used to accomplish the decision tree analysis.

Before describing how patterns characterising different classes are identified and expressed first as decision trees and then as rules, a brief

description of the terminology associated with decision trees is given below.

Describing a decision tree is analogous to describing a tree in nature. The point where the tree begins is the root (see Figure 3.3). Arising from the root is a trunk that gives off branches. A branch is said to end in a leaf when no further branches arise from it. If further branches arise from it, the point at which they do is called a node.

Figure 3.3 (adapted from Quinlan, 1993) shows a fragment of a tree that identifies the patterns of weather conditions most frequently associated with days when the weather is conducive to playing golf. Note that in contrast to their biological counterpart, it is common practice to draw decision trees with the root on top.

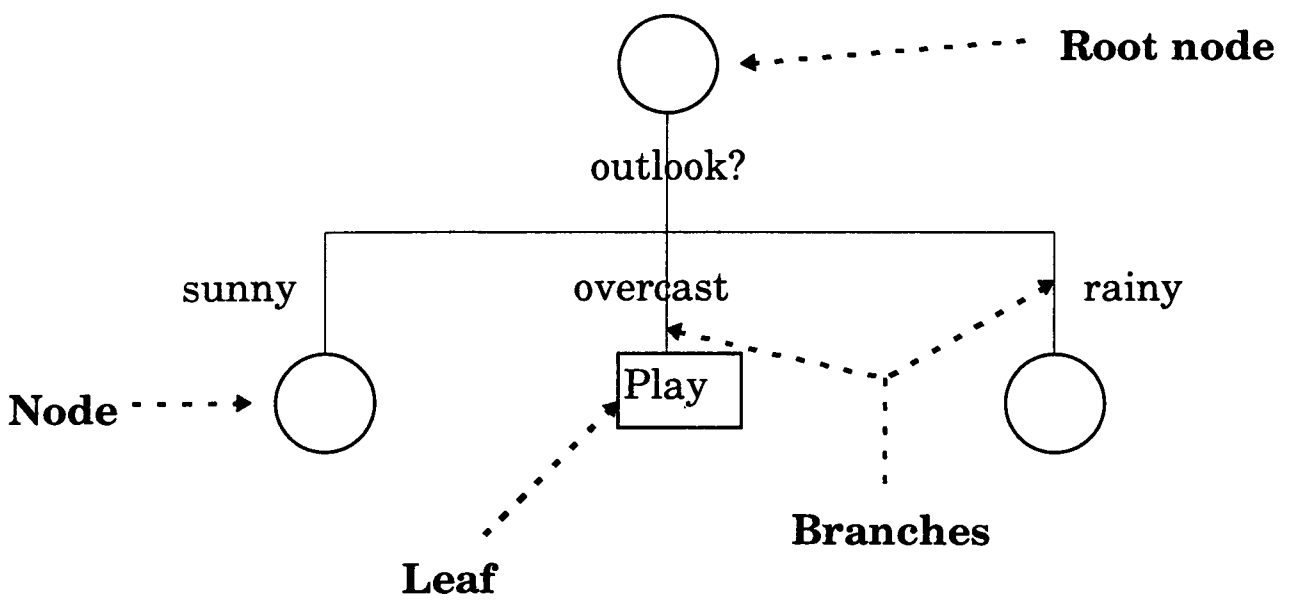


Figure 3.3 A simple decision tree

3.5.2 Data used to construct trees and rules

Full details of the construction of decision trees and rules are beyond the scope of the thesis. This section and those following it therefore, briefly describe those aspects that are essential for understanding the rest of this thesis. More information can be found in Quinlan's book, 'C4.5: programs for machine learning', 1993.

To identify the pattern of features and express this as a decision tree, the data has to be presented to a program such as C4.5 as a set of **cases**, called a **training set.** Cases in the set are pre-classified as belonging to different **classes**. Each case is described by a number of features. For example, Table 3.6 shows a set of 14 cases for constructing a tree such as that in Figure 3.4. Each case is described by 4 features, namely - outlook, temperature, humidity and wind and represents a day when the weather conditions rendered it conducive or not conducive to play golf.

Table 3.6 Training cases for constructing a tree

Case No.	Outlook	Temp (°F)	Humidity (%)	Windy	Class
1	sunny	75	70	true	play
2	sunny	80	90	true	don't play
3	sunny	85	85	false	don't play
4	sunny	72	95	false	don't play
5	sunny	69	70	false	play
6	overcast	72	90	true	play
7	overcast	83	78	false	play
8	overcast	64	65	true	play
9	overcast	81	75	false	play
10	rainy	71	80	true	don't play
11	rainy	65	70	true	don't play

12	rainy	75	80	false	play
13	rainy	68	80	false	play
14	rainy	70	96	false	play

3.5.3 Extracting feature combinations to build trees and rules

From the cases available C4.5 extracts the combination of features (i.e. the weather conditions) that render the day conducive or not conducive to playing golf. The patterns are first expressed as decision trees and then converted to rules. The decision tree is built methodically by C4.5 in the steps described below. Combining the reading of the text with references to Table 3.6 and Figure 3.4 should aid an easy understanding.

1. The tree first starts as a root node containing all the training cases. C4.5 looks to see if all cases at this node belong to the same class. If they don't, as in the example here, it then grows branches from this node using the 'best' feature that sorts the cases into distinct groups based on the different values of the feature. The test that determines which feature is the 'best' considers the ability of the feature to sort the cases into classes, i.e. when most cases with the same feature value also have the same class value. In the example here, outlook is chosen as the 'best' feature to split the root node. It has 3 values associated with it, sunny, overcast and rainy. This feature is obviously quite useful for sorting the cases into classes, as the middle group, overcast contains only cases belonging to the class 'play'. The first and third, i.e. sunny and rainy, subsets, however still have mixed classes.

2. Following the split at the root node, the resulting nodes are next examined. If all the cases in a node belong to the same class, then this node is declared a leaf and the class of the cases is assigned to it. In the example here, as the node corresponding to the branch of the middle group, overcast, only has cases of class play, this node is labelled as the leaf 'play'.

3. If all the cases in a node do not belong to the same class (as here with nodes corresponding to the branches sunny and rainy), C4.5 grows further branches at this node to sort the cases apart. The same procedure as in Step 1. is used, i.e., at each node the feature that can best sort the cases is chosen to grow the branches. In the example here, the first subset, sunny is further divided by the feature humidity with outcomes, humidity less than or equal to 75 and humidity greater than 75. The third subset, rainy is further divided by the feature, windy with outcomes, windy or not windy. Each subset now contains cases from a single class. The nodes are therefore declared as leaves. As there are no further nodes left to expand, the procedure is completed.

The final tree from this procedure is shown in Figure 3.4

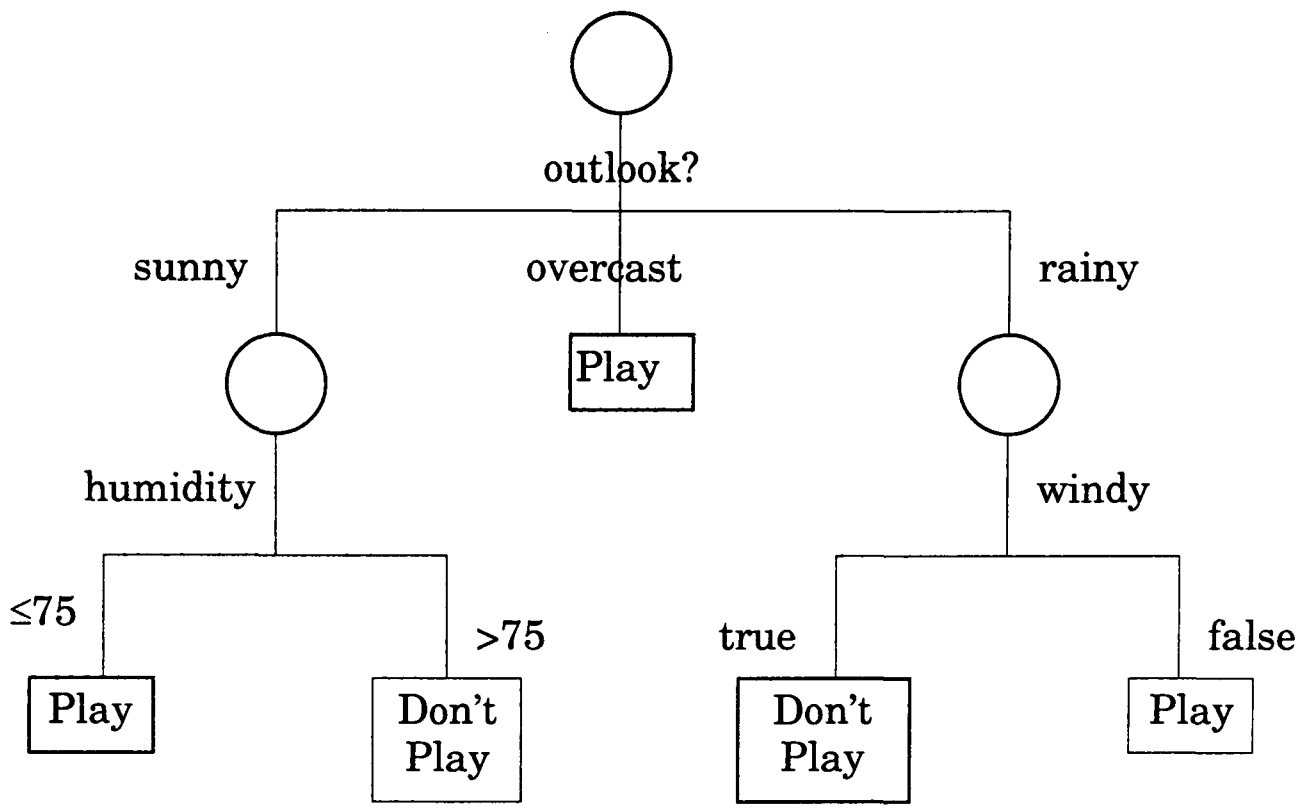


Figure 3.4 To play or not to play: the complete decision tree

Converting trees to rules

For ease of comprehensibility, patterns expressed as trees are converted to a set of rules. Every path traced from the root of a tree to a leaf gives a rule. For example, here are the rules obtained by converting the left subtree in Figure 3.4:

Rule 1

If
the outlook is sunny and
the humidity less than or equal to 75
Then
Class = **play**

Rule 2

If
the outlook is sunny and
the humidity is more than 75
Then
Class = **don't play**

C4.5 excludes some rules that are superfluous or whose exclusion is not expected to lower the predictive accuracy (explained in Section 3.5.4 below). This may leave some cases unclassified by any rule. To cover for this, the conversion to rules always introduces a 'default' rule, which specifies how such cases are to be classified.

3.5.4 What is the accuracy of the rules?

Apparent accuracy

Having identified rules characterising the classes in the training set, it is desirable to determine the accuracy of the rule set. Apparent accuracy is the ability of the rule set to assign correctly cases in the training set to their classes. This would indicate how effectively the combinations of features or patterns characterising the different classes have been captured in the rule set. It is calculated by using the rule set to classify each case in the training set and comparing this classification against the real class that the case belongs to. This can be best summarised by drawing a contingency, or 2×2 table. For the example given here, when the rule set obtained from the tree in Figure 3.4 is used to assign classes to the cases in the training set, this would appear as follows:

		Actual class		
		Play	Don't Play	
Predicted class	Play	Tp (9)	Fp (0)	N3 (9)
	Don't Play	Fn (0)	Tn (5)	N4 (5)
		N1 (9)	N2 (5)	N (14)

'Predicted class' is the class assigned to a case by the rule it satisfies. The entries in this table have the following meaning. Tp, or true positives, are the number of cases (here 9) that belong to class 'play' and are classified as 'play'. Fp, or false positives, are the number of cases (here 0) that belong to the class 'don't play' that are incorrectly classified as 'play'. Similarly, Fn, or false negatives, are the number of 'play' cases (here 0) incorrectly classified as 'don't play' and Tn, or true negatives, are the number of 'don't play' cases (here 5) correctly classified as such. N1 is the total number of cases that belong to the class 'play' (here 9), calculated by adding Tp and Fn. N2 is the total number of cases that belong to the class 'don't play' (here 5), calculated by adding Fp and Tn. Similarly, N3 is the total number of cases classified by the rule set as 'play' (here 9), calculated by adding Tp and Fp. N4 is the total number of cases classified by the rule set as 'don't play' (here 5), calculated by adding Fn and Tn. N gives the total number of cases in the training set. The accuracy of the rule set is calculated as the fraction of cases correctly classified. That is:

(Tp + Tn)/N or in the example here (9 + 5)/14 = 1

This fraction is sometimes expressed as a percentage by multiplying the number obtained by 100. The apparent accuracy in the example here therefore is 1 or 100%, that is, the rules were able to classify all the training cases accurately.

Predictive accuracy

It is not sufficient to know just the apparent accuracy of a rule set. It is common for rules to have high apparent accuracy. The question however is, how does this rule set perform on new data? The accuracy of classification of *new* data by the rule set extracted from the training set is called the predictive accuracy. This performance on the new data indicates whether the patterns identified on the training data can be generalised. Importance can however be attached to the predictive accuracy value, only if the classification of new data using the rule set, as summarised by the corresponding 2×2 table, is significantly different from a classification obtainable by chance (e.g. tossing a coin) on the same data. This is assessed by performing a chi-square (χ^2) test on the 2×2 table for the new data. For details on this procedure, see Everitt (1992). For the purpose of this thesis, it is sufficient to know that a χ^2 value > 3.84 indicates that there is a less than 5% probability of obtaining this classification by chance ($P<0.05$). A high χ^2 value together with a high predictive accuracy, therefore suggests that some intrinsic principles concerning the embryo groups have been captured by the rule set.

Figure 3.5 below, recapitulates the steps described above undertaken to determine whether the patterns identified in the training sample hold good on new data and can therefore be generalised.

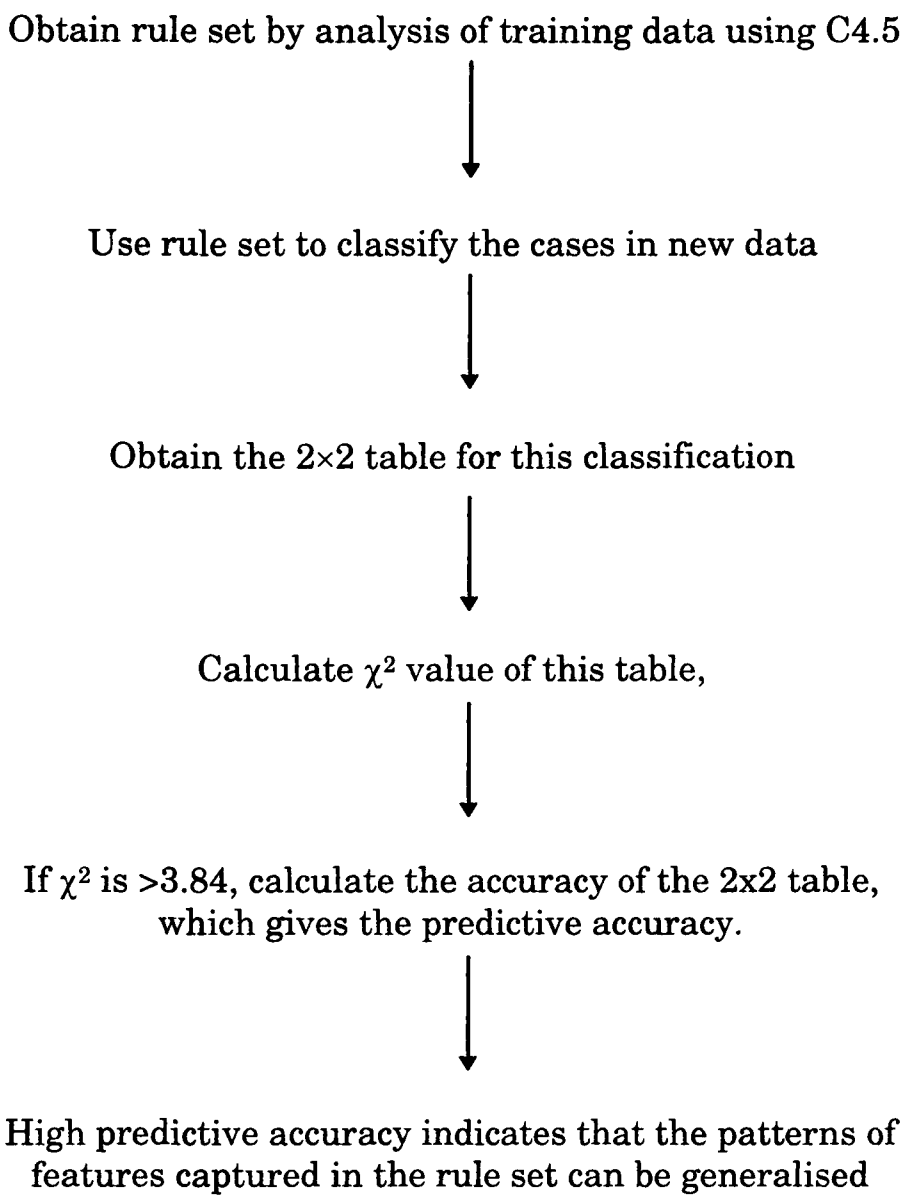


Figure 3.5 Can the patterns be generalised?

If new data is not available, it is still possible to obtain an idea (or estimate) of the predictive accuracy that can generally be expected. As the number of training cases increases, the apparent accuracy rate comes closer to the predictive accuracy. In the real world however, sample sizes are usually quite modest. Certain techniques are therefore used to obtain an estimate of the predictive accuracy. The estimate obtained by these

techniques (described below) is fairly accurate and the predictive accuracy of the rule set on new data will usually be within 2 standard deviations of the estimated predictive accuracy.

Estimating predictive accuracy

The Cross-validation or leave 'k' out technique which is used when the number of training cases is small, typically less than 200, was used in this dissertation to obtain estimates of the predictive accuracy (Weiss and Kulkowski, 1991). Here the sample (N= total number of cases in the sample) is divided so that a few cases (k) are 'left out' and used as the 'test' set. The 'test' set serves the role of new data. The rules obtained by training on the remaining cases (N-k) are tested for their classification ability on the test set. This procedure is repeated by 'leaving out' a different group of k cases each time selected at random. Thus each case in the sample is used as a test case and each time most of the cases are used for training. The estimate of the predictive accuracy of the rule set obtained by training on all the cases is then the average of the predictive accuracy for each test sample.

A special case is the procedure called 'leave one out' where k=1. Since only one case is left out each time, the procedure is repeated a large number of times. The estimate of the predictive accuracy by the 'leave one out' procedure is an almost unbiased estimation of the predictive accuracy. This is used when the sample size is very small (less than 100).

It is possible to construct a single 2×2 table for the result of a cross-validation. The entries of this table are obtained by adding corresponding cells from each of the 2×2 tables obtained on the k cases left out in turn. χ^2 values and predictive accuracy estimates can be obtained from this table.

For example, the 'leave one out' procedure was used to obtain an estimate of the predictive accuracy of the tree in Figure 3.4. The 2×2 table shown below was obtained by adding the corresponding cells for each of the 14 2×2 tables obtained on the single case left out each time

		Actual class		
		Play	Dont Play	
Predicted class	Play	5	3	8
	Dont Play	4	2	6
		9	5	14

The χ^2 value of this table is 0.21 (as this is <3.84, the hypothesis that this classification could be by chance, cannot be rejected). An estimate of the predictive accuracy is 5+2/14 = 0.5 or 50%.

This performance illustrates how the apparent accuracy rate on the training set (here 100%) can be quite misleading, when the rules are obtained from very small training sets.

3.5.5 Simplification of trees and rules

Sometimes, the tree (and therefore the rule set) obtained may be very large and complicated ending in a leaf for almost every training case. This pattern is thus highly specific to the training cases and is said to 'overfit'. This indicates that the pattern may not hold good on new data. This is illustrated by the fact that when such a tree is used to classify some new cases, many of these may be classified incorrectly when they are even slightly different from the cases in the training set. The predictive accuracy will therefore typically be very low. Trees therefore need to be simplified (i.e. decrease in the size), so that a leaf can accommodate more than just a specific case. Beyond a certain level of simplicity however, further shortening or simplification of the tree may prove counterproductive as the rule set is so general that cases belonging to one particular class are easily accommodated by a rule set that is supposed to characterise another class. Cases could therefore be misclassified, again resulting in a lowering of the predictive accuracy.

C4.5 uses 2 parameters to control the complexity of the trees obtained. These parameters are called 'm' and 'c'. Different 'm' values indicate the minimum number of cases that should be allowed in a 'leaf'. A higher 'm' value therefore results in a smaller tree as the further formation of subtrees at a node (with 'm' number of cases) is prevented. The node is converted to a leaf, and given the class to which a majority of the cases at the node belong.

A tree can also be simplified by discarding one or more sub-trees and replacing them with leaves. This process is controlled by the 'c' setting. A higher 'c' value results in a bigger tree as fewer sub-trees are allowed to be replaced by leaves. As 'm' and 'c' are varied, we would typically expect to see the following relationship between the size of a tree and its predictive accuracy (see Figure 3.6).

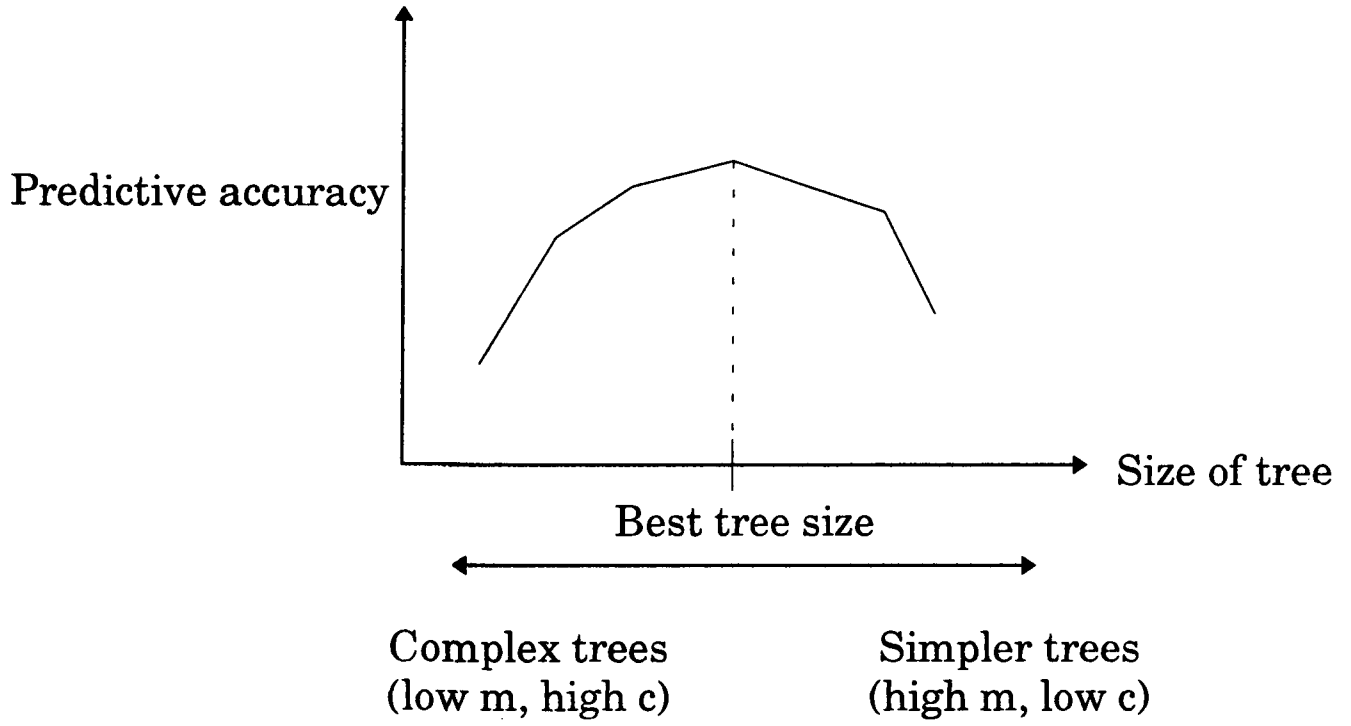


Figure 3.6 Altering 'm' and 'c' settings to obtain the tree with the highest predictive accuracy

Decision tree construction with C4.5 therefore involves altering 'm' and 'c' settings in an attempt to identify the tree with the highest predictive accuracy. Section 3.6 below describes further the procedure used in this dissertation to find this best sized tree.

3.6 Constructing trees and rules in this dissertation

Chapter 4 investigates the feasibility of using the decision tree techniques described in this chapter, to analyse IVF data. In chapter 5, the techniques are used to identify the combinations of features (i.e. the patterns) characterising embryos of differing developmental potential as determined by the outcome of transfer. The results are then discussed with a view to understanding the relationship between embryo-related features comprising the pattern and the developmental potential of embryos. Chapter 6 shows how the techniques can be extended to investigate other problems in IVF.

The following procedure is used in the dissertation to obtain rules that perform significantly better than chance and have a high predictive accuracy:

1. Embryos of known developmental potential (for example, 'pregnant' or 'not pregnant' on transfer) are identified and grouped.
2. A number of features related to each embryo in each group or class are collected from the records.
3. The data are analysed using C4.5 and a set of trees from complex to simple is constructed (using settings for C4.5 parameters). For each tree:

    a) a C4.5 rule set is obtained

b) Due to the size of the samples, the method used to obtain an estimate of the performance of the rule set on new data is cross validation. Where the sample number is < 100 (Chapters 4 and 6). a 'leave one out' procedure is used. Where the sample size is between 100 and 200 (Chapter 5), a 'leave 10 out' procedure is used. These steps are in accordance with Weiss and Kulkowski, 1991 and are implemented here using an automated procedure[9].

Only those rule sets whose cross validation 2×2 table has a χ^2 value > 3.84 are considered further. The estimated predictive accuracy for these rule sets is obtained.

4. The 'best'[10] rule set is the one that has the highest estimated predictive accuracy[11].

5. The relationship between the features contained in the rule set and the developmental potential is examined.

---

[9]Computer program written by Dr Ashwin Srinivasan of the Oxford University Computing Laboratory.

[10]The term 'best' is used with the following qualification. In theory ,values of 'm' can vary from 1 to the number of cases available. Value of 'c' can similarly range from 1-100. In practice however, a tractable search for 'm' and 'c' settings is forced to look at far fewer values. This is done either by restricting the range and/or considering only some values within the specified range for 'm' and 'c' values. Some experimentation on the given data is required to decide on these restrictions. This results in a manageable number of settings being tried in step 3.

[11] In case of rule sets with equal estimated predictive accuracy, the rule set with fewer rules was selected as it would be simpler while at the same time capturing the essential pattern. This is in line with 'Occham's razor', the general principle advocated in many areas in science, which states that : All other things being equal, simpler theories are to be preferred.

3.7 Summary

This chapter has outlined the materials used and the methodology followed in the Oxford IVF Unit, and details of the technique of decision trees used in this dissertation. Particular emphasis has been placed on the features observed and recorded in the treatment files of the IVF patients. Against this backdrop, the next chapter describes a pilot study that investigates the application of decision tree techniques to IVF data.

Chapter Four. Pilot Study

4.1 INTRODUCTION 81

4.2 ORGANISATION OF THIS CHAPTER 81

4.3 MATERIALS AND METHODS 82

4.3.1 Feature Selection. 82

4.3.2 Embryo Selection 83

4.3.3 Data Collection And Conversion 88

4.3.4 Obtaining Rules 92

4.4 RESULTS 93

4.5 DISCUSSION AND CONCLUSIONS 95

4.6 SUMMARY 97

4. PILOT STUDY

4.1 Introduction

This chapter presents a feasibility study of the use of decision tree techniques to analyse IVF data. The amenability of such data to conversion into a form acceptable by decision trees, analysis using the decision tree program C4.5 and the expression of results in the form of simple rules are explored. The analysis described in this chapter includes 81 embryos of known subsequent histories. 47 embryos are classified as 'good' and 34 as 'bad' based on the successful clinical pregnancy outcome after transfer *in vivo* or the levels of secretion of human chorionic gonadotrophin (hCG) on *in vitro* culture. 15 features related to each of these 81 embryos were extracted from IVF treatment files and comprise the data used for this pilot study. The ability of decision trees to analyse this data and express the relationship between the 15 features and the 'good' or 'bad' developmental potential of an embryo in the form of easily interpretable rules was investigated and is presented here.

4.2 Organisation of this chapter

The Materials and Methods section (Section 4.3) below, details the following: (a) the 15 features selected; (b) selection and classification of embryos; (c) collection of data from patient files; (d) conversion of data to a form ready for analysis by C4.5; and (e) the procedure of obtaining the rules. The rules obtained are presented in the Results section (Section

4.4). The feasibility of the use of C4.5 is assessed in the Discussion and Conclusions section (Section 4.5), on the basis of rules obtained and the ease of their interpretation along biological lines.

4.3 Materials and methods

Figure 4.1 gives in broad outline, the steps followed in this study. Each step is elaborated further in relevant sections below.

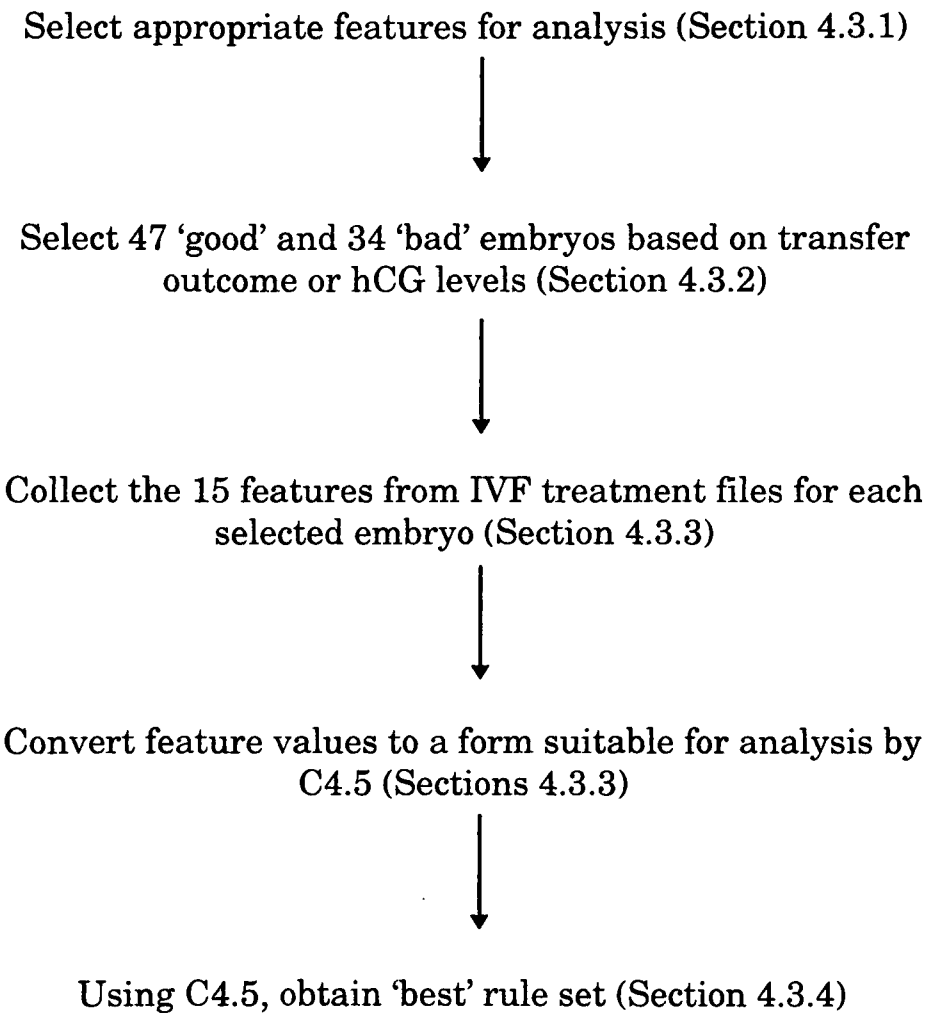


Figure 4.1 Outline of the materials and methods

4.3.1 Feature selection.

A number of features relating to the oocyte, follicle, sperm and embryo as well as the patient and the partner (or donor) are recorded by the IVF

Unit staff in the patient notes. The features to be used in the study were selected following discussions with the embryologists and observation of the procedures in the IVF clinic. As explained in Chapter 1, in the interest of a simple experimental design, only embryo-related features (i.e. features describing the follicle, oocyte and embryo) were considered. These features could be traced to individual embryos, for example cell number, or size of follicle of origin of the embryo. Of the features considered, the inclusion process was biased to those that were commonly used by the embryologists during the selection procedure. Thus a total of 15 embryo-related features (see Table 4.1) was selected. Of these, 8 described the embryo, 6 were related to the oocyte of origin and 1 related to the follicle of origin of the oocyte.

4.3.2 Embryo selection

Embryos with a known subsequent developmental outcome were selected to allow classification into 'good' and 'bad' classes.

Figure 4.2 identifies the categories of embryos included in the 2 classes. The rationale underlying the selection of these embryos is explained, below.

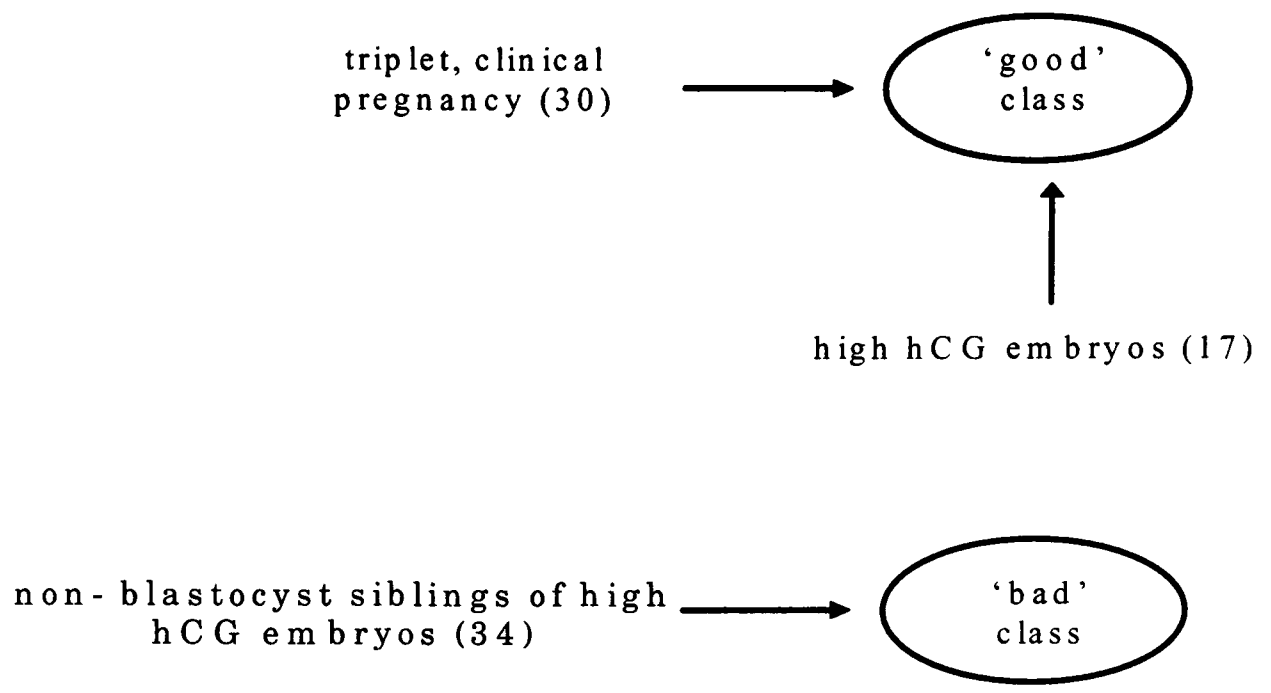


Figure 4.2 Categories of embryos included in the 'good' and 'bad' classes

'Good' embryos

Embryos included here were selected from the following categories:

1. **Triplet pregnancies:** A maximum of 3 embryos are selected for transfer to patients in the Oxford IVF Unit (see Section 3.4.5). The patients treated in the Unit during January 1991 to December 1992 who progressed after 3 embryo transfer to a triplet clinical pregnancy (10 patients) diagnosed by an ultrasound scan at 6 weeks after transfer, were identified. Such a diagnosis implied the 'good' developmental potential of embryos transferred. Data for these embryos, extracted from treatment records, were therefore included in the 'good' class

2. **High hCG embryos:** 31 spare embryos (see Section 3.4.7) cultured *in vitro* in T6 medium supplemented with 10% autologous maternal serum and progressing to the blastocyst stage of development were considered. The levels of hCG measured daily in the ECF of individual embryos, from day 3 to day 14 of culture were available from a previous study (Dokras *et al*, 1993).

   hCG is a secretory product of the early conceptus and levels are considered indicative of the well being of an embryo. In the *in vivo* situation, biochemical pregnancies, ectopics and spontaneous abortions show lower serum hCG levels than normal singleton pregnancy (Confino *et al* 1986). *In vitro*, a correlation between levels of hCG detected in ECF and blastocyst quality as assessed morphologically, has also been demonstrated (Dokras *et al*, 1993).

   Overall, peak levels of hCG were detected on day 10 of culture for the 31 embryos considered here. After day 10, levels showed a decline, possibly due to the inability of culture medium to sustain further development. In this study therefore, levels of hCG secreted up to day 10 were considered and cumulative hCG levels detected from day 3-day 10 of culture calculated (see histogram in Figure 4.3).

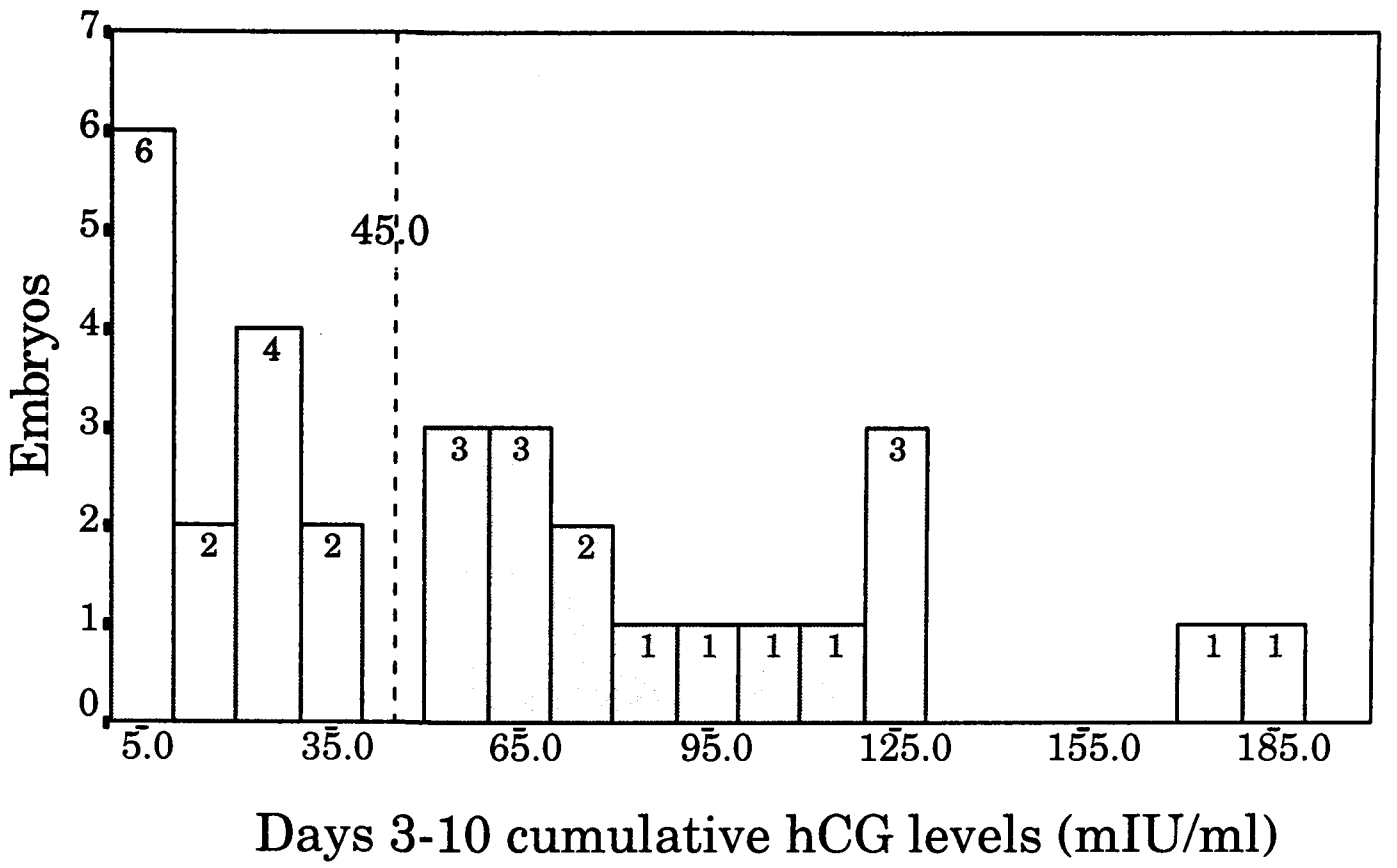


Figure 4.3 Histogram showing the distribution of embryos according to cumulative (days 3-10) hCG values

A clustering of 14 embryos secreting <45mIU/ml of hCG and their delineation from those secreting higher levels is visible. Although 45mIU was arbitrarily chosen as the cut-off value, it was found to correspond to differences in embryo quality as defined by the grading of Dokras *et al*, 1993 (details of grading in Section 6.3.1). Low hCG levels were found to be associated with poor quality (blastocyst Grade 3) and high hCG levels with good quality (blastocyst Grade 1 and 2) respectively. The 17 'high' (>45mIU/ml) hCG secretors were therefore included in this study in the 'good' embryo group. The features associated with these embryos at the earlier stages of development were extracted from the records.

'Bad' embryos

Embryos to be included in this group were not self-evident. The following categories of embryos were initially considered:

1. Embryos not resulting in pregnancy on transfer.
2. Embryos progressing on culture to low hCG secreting blastocysts (see Figure 4.3) *in vitro.*
3. Non-blastocyst forming siblings of the high hCG secreting blastocysts.

Of these, candidates in categories 1 and 2 did not qualify for the following reasons. In the case of category 1, if transferred embryos did not result in a pregnancy, it was not immediately obvious that the cause for failure was the 'poor' or 'bad' quality of the embryos transferred. A hostile maternal *milieu* could have been responsible. In situations where embryo transfer resulted in a singleton or twin (rather than triplet) clinical pregnancy, thus demonstrating a uterine environment conducive to implantation, it was not possible to determine which of the 3 embryos were 'bad' and had failed to implant. Identification and inclusion of 'bad' embryos from amongst those that had been transferred was therefore not possible. Similarly, candidates in category 2 belied inclusion in the 'bad' group. The incubator conditions and T6 medium used for culturing these embryos, were similar to those used for the embryos that resulted in 'high hCG' secreting blastocysts. However the maternal serum supplementing the culture medium varied for each patient's embryos. Deficiency of the culture conditions rather than any inherent defect in the embryo could

therefore be responsible for the secretion of 'low' levels of hCG and this possibility precluded the classification of these embryos as 'bad'.

Embryos in category 3 were included in the 'bad' group. These were embryos that did not progress even as far as the blastocyst stage although *in vitro* culture conditions were favourable; deduced from the fact that siblings of these embryos when cultured under similar conditions (same maternal serum supplement) progressed to become 'high' hCG secreting blastocysts. This therefore pointed towards an intrinsic defect in the non-blastocyst forming embryos thus justifying their inclusion in the 'bad' group.

4.3.3 Data collection and conversion

Data relating to the 15 selected features for each of the 81 embryos was collected after discussions with the embryologists. These discussions gave a better understanding of the method of recording followed in the IVF clinic, and also aided in deciphering the abbreviations used when recording features in the files. Some features were recorded as simple 'yes' or 'no' values e.g. visibility of nuclei in the cells; others were recorded as qualitative descriptions, e.g. cell regularity recorded as regular, slightly irregular, irregular, very irregular (all written as abbreviations); and finally some features were recorded as numerical values, e.g. follicle size = 16.5mm, 18.75mm, etc. For the purpose of

analysing the data using the decision tree program, the recorded values for different features were assigned numbers or codes.

Depending on the properties of the feature, the assigned numbers or codes were either nominal, ordinal or ratio measurements (SPSS, 1993). Briefly, *nominal* values are simply labels for different categories. Neither category is higher nor lower than the other. (e.g. here visibility of nuclei is expressed as three categories 'yes', 'no' and 'not applicable'). *Ordinals* are ordered without notions of distances between values. They provide information about whether a feature value is higher or lower than another but cannot say by how much (e.g. cell regularity). Thus, on a ordinal scale measuring cell irregularity, 'regular' < 'slightly irregular' < 'irregular' < 'very irregular'. In order to enable the decision tree program used here to compare ordinal values, the labels were given numeric codes in consultation with the embryologist, to reflect their position in the corresponding ordering. Thus, the numeric codes for cell regularity are: 'regular' = 1, 'slightly irregular' = 2, 'irregular' = 3, 'very irregular' = 4. It must be reiterated that these numbers, in case of ordinals are only used for comparison, and no attempt should made to interpret any 'distance' based relationships amongst them. Thus, the statement that a cell that is 'very irregular' (code = 4) is twice as irregular as one that is 'slightly irregular' (code = 2) is meaningless. *Ratio* values are ordered with distances between values being well defined; e.g. follicle size values

14mm, 15mm, 16mm. Such values not only give information about which value is higher or lower, but also by how much.

The features used for analysis were described in detail in Chapter 3 (see Section 3.4). For completeness, a full list of the 15 features along with the level of measurement and their assigned numbers or codes is shown in Table 4.1

Table 4.1 15 features selected for analysis

No	Feature	Range of values	Level	Number or code
1	follicle size recorded at oocyte recovery (day 0)	10 - 38 mm	ratio	same as original value[1]
2	oocyte cumulus amount (day 0)	very little amount medium lots	ordinal	1 2 3
3	oocyte cumulus description (day 0)	dark (immature)[2] slightly dark dark (just post mature)[2] very slightly post mature slightly post mature post mature	ordinal	1 1 1 2 3 4
4	oocyte corona maturity (day 0)	very immature immature mature	ordinal	1 2 3

[1] In some cases, the size of the follicle of origin could not be ascertained either due to the mistaken aspiration of more than one follicle, the aspiration of fluid from a ruptured follicle collected in the Pouch of Douglas or re-aspiration of a follicle that originally yielded no oocyte but subsequently re-expanded. Following Quinlan, 1993, the value was denoted by '?'.

[2]The IVF records use the same keyword "dark" in two different contexts. It is impossible to distinguish between darkness related to immaturity, and that related to postmaturity. To simplify this, all dark values are given the same numerical code.

5	oocyte corona description (day 0)	okay	ordinal	1
		slightly dark		2
		dark		3
		borderline		4
		very dark		5
6	egg grade	post mature	ordinal	1
		mature		2
		borderline immature		3
		immature		4
7	cumulus dish attachment	massive	ordinal	1
		lots		2
		moderate		3
		little		4
		none		5
8	number of cells (check 1, day 2)	1 - 10	ratio	same
9	regularity of cells (check 1, day 2)	not applicable[3]	ordinal	?
		regular		1
		slightly irregular		2
		irregular		3
		very irregular		4
10	fragmentation (check 1, day 2)	no fragments	ordinal	1
		few tiny fragments		2
		few small fragments		3
		few fragments		4
		fragments		5
		lots of fragments		6
11	visibility of nuclei in the cells (check 1, day 2)	not applicable[4]	nominal	?
		no		no
		yes		yes
12	number of cells (check 2, day 2)	1 - 10	ratio	same
13	regularity of cells (check 2, day 2)	not applicable	ordinal	?
		regular		1
		slightly irregular		2
		irregular		3
		very irregular		4

[3]Regularity across cells and visibility of nuclei are not applicable at the pronuclear and syngamy stage. Following Quinlan, 1993, values that are not applicable are denoted by '?'

[4] See explanation in footnote 3

14	fragmentation (check 2, day 2)	no fragments	ordinal	1
		few tiny fragments		2
		few small fragments		3
		few fragments		4
		fragments		5
		lots of fragments		6
15	visibility of nuclei in the cells (check 2, day 2)	not applicable	nominal	?
		no		no
		yes		yes

Table 4.2 shows the first few items of data in the final form ready to be used as the input for C4.5.

Table 4.2 Examples of data in the final form ready for analysis

	Value of feature for each embryo				
Case No.	Follicle size	cell number (check 1 day 2)	Cell regularity (check 1 day 2)	...	Class
Embryo 1	14.00	2	3	...	bad
Embryo 2	14.45	3	4	...	bad
Embryo 3	18.00	4	1	...	good
Embryo 4	16.67	5	4	...	bad
...	...	...	...	...	...
Embryo 'n'	19.25	4.0	2	...	good

4.3.4 Obtaining rules

The data for the 81 embryos were used by C4.5 to construct the 'best' rule set i.e. the rule set with the highest estimated predictive accuracy (see Section 3.5.4). A total of 1330 settings (with different combinations of 'm' and 'c' parameters) were tried, giving 1330 rule sets. The predictive

accuracy for each of these 1330 rules sets was estimated by the 'leave one out' procedure (see Section 3.5.4).

Two distinct rule sets with the same 'highest' predictive accuracy estimate were obtained. The rule set that was smaller was selected and is presented in the Results Section.

4.4 Results

First the rules obtained are shown. Then the 2×2 table representing the performance from 'leave one out' cross validation, its χ^2 value and estimate of predictive accuracy are given.

Rules

Rule 1

If

cell count (check 2, day 2) > 3 and
cell count (check 2, day 2) ≤ 4

Then

class = **good**

Rule 2

If

cell count (check 1, day 2) > 3 and
fragments (check 1, day 2) ≤ 4 and
visibility of nuclei (check 2, day 2) = no

Then,

Class = **good**

Rule 3

If

regularity (check 1, day 2) ≤ 2 and
regularity (check 2, day 2) > 1 and
follicle size > 15

Then

Class = **good**

Rule 4

If

cell count (check 2, day 2) ≤ 3

Then

Class = **bad**

Rule 5

If

cell count (check 2, day 2) > 4 and
visibility of nuclei (check 2, day 2)= yes

Then

Class = **bad**

Rule 6

If

fragments (check 1, day 2) > 4

Then

Class = **bad**

Rule 7 (default rule)

If
none of the above

Then

Class = **good**

'leave 1 out' cross validation 2×2 table

		Actual class		
		good	bad	
Predicted class	good	43	13	56
	bad	4	21	25
		47	34	81

χ^2 value = 23.8

Estimate of predictive accuracy of the rule set = 79% (±standard deviation 4.5)

4.5 Discussion and Conclusions

The feasibility of using the decision tree technique to analyse IVF data was investigated. Data relating to 15 features collected for 47 'good' and 34 'bad' embryos were analysed.

The collection procedure, the analysis and results obtained indicate the following:

1. Extraction of data related to embryos, from the patient files was possible after thorough consultations with embryologists.

2. The data, after some processing submits itself easily to analysis by C4.5.

3. Data related to a large number of features is easily handled and as can be seen from the results (see Section 4.4), interdependencies amongst features are explored and expressed. For example, consider rule 2 which gives the combination of values of the number of cells, fragments and nuclear visibility that must be satisfied if an embryo is to have a 'good' developmental potential.

4. The high χ^2 value and the high estimate of predictive accuracy (79%) indicate that the classification in the 2×2 table was not obtained by chance and that an inherent pattern characterising the two groups was captured by the rules.

5. The results are stated as simple easily understandable rules and they allow a smooth interpretation along biological lines. For example,

consider rule 1 (Section 4.4). This rule has 2 conditions which taken together indicate that an embryo at the time of the second check on the day of transfer showing more than 3 but less than or equal to 4 cells, has a 'good' developmental potential. In other words, the rule conveys that a cleavage rate that is neither too slow nor too fast is ideal. The rule is thus easily interpretable.

Similarly consider rule 2 which is slightly more complex.. The component conditions of this rule indicate the following (a) If an embryo at the time of the first check on the day of transfer has more than 3 cells; and (b) at the same check is found to have few or tiny or small fragments or no fragments (see feature 10 in Table 4.1 for decoding of value 4); and (c) at the time of the second check on the day of transfer shows no visible nuclei within cells, then, the embryo has a 'good' developmental potential. As with rule 1, this rule offers no difficulty in the interpretation of the features and associated values. Similarly, the other rules lend themselves to an easy interpretation.

This study however included data from *in vitro* cultured as well as *in vivo* transferred embryos, the development of each group being influenced by different factors. Looking towards the results obtained here, with a view to gaining any biological insight into the relationship between features and development would therefore be outside the modest scope of this study.

4.6 Summary

This chapter explored the ability of C4.5 to analyse IVF data. Data related to 15 features for 81 embryos were investigated. The results obtained were accurate and easily understandable. The role of the study in this chapter was only to demonstrate the applicability of a particular form of analysis. The next Chapter describes the actual use of this analysis method in investigating the relationships between embryo, oocyte and follicle features and the developmental potential of embryos. Data relating to 53 features for each of 600 embryos transferred to patients is analysed in Chapter 5 with the pregnancy outcome serving as the scale for assessing developmental potential.

Chapter Five. Batch Study

5.1 INTRODUCTION 99

5.2 ORGANISATION OF THIS CHAPTER 101

5.3 MATERIALS AND METHODS 102

5.3.1 Feature Selection. 102

5.3.2 Patient Selection 103

5.3.3 Batch Selection And Classification 105

5.3.4 Data Collection And Conversion 105

5.3.5 Batch Average 106

5.3.6 Obtaining Rules 107

5.3.7 What Does 'Embryo Grade' Capture? 110

5.4 SCOPE OF THE STUDY 111

5.5 RESULTS 114

5.5.1 Results Obtained Using All 53 Features 115

5.5.2 Results Obtained Excluding Embryo Grade 117

5.5.3 Further Analysis 119

5.6 DISCUSSION 125

5.6.1 Rules Obtained Using All Features (Rule Set 1) 125

5.6.2 What Does 'Embryo Grade' Capture (Rule Set 2)? 129

5.6.3 Relevance To Current Assessment Procedures 131

5.6.4 Shortcomings Of The Experimental Design Of The Batch Study 134

5.7 A NOTE ON USING LOGISTIC REGRESSION 136

5.7.1 Relevance To Results Obtained With Decision Trees 140

5.8 CONCLUSION 141

5.9 SUMMARY 141

5. BATCH STUDY

5.1 Introduction

The preliminary study in Chapter 4 demonstrated the ability of the decision tree technique to analyse IVF data and express the results as simple easily understandable rules. Given the promising nature of these results, this chapter reports on the use of decision tree techniques for gaining a more clear and accurate understanding of the relationships between embryo-related features and the developmental potential. As explained in Section 4.5, such insight cannot be obtained from the pilot study which uses a combination of *in vivo* transferred and *in vitro* cultured embryos and is a small study involving only 81 embryos. The study reported in this chapter, uses data for embryos transferred, in batches of three, to IVF patients. Pregnancy outcome from the transfer is used as the criterion to assess the development. Features describing the transferred embryos were extracted from the IVF treatment files of patients and translated into a form suitable for investigating the pattern of features characterising embryos that differ in their transfer outcome. Identification of the important patterns could assist in a more informed selection of embryos for transfer during *in vitro* fertilisation treatment with the possibility of an increase in IVF success rates.

The salient characteristics of the batch study are as follows:

1. **More features:** The IVF Unit in Oxford has detailed records for a large number of embryo, oocyte, sperm, follicle, maternal and paternal features recorded by the embryologists and other trained staff. The batch study takes advantage of this information and uses 53 embryo-related features in the analysis (details in Section 5.3.1)

2. **Batches of embryos:** Batches of transferred embryos (three embryos constitute a batch) are used as the unit for analysis, unlike the pilot study which used individual embryos. A batch resulting in the birth of a singleton or twins is labelled a 'take-home baby (TH)' batch as is a batch resulting in triplets. Similarly a batch not resulting in a pregnancy is labelled a 'no take home baby (NTH)' batch. The risk here, as indicated by the singleton and twin births, is that all embryos comprising the 'TH' batch may not essentially be embryos capable of resulting in a 'take home' baby. Similarly, not all embryos in a 'NTH' batch may be incapable of resulting in a pregnancy. In some cases, maternal factors rather than the embryos could be responsible for the failure. The use of batches labelled 'TH' and 'NTH' instead of individual embryos of known outcome thus carries with it the risk of 'weakening' the data. The advantage however, is that this largely increases the availability of data for analysis. The canvas of choice extends to all patients who have had IVF treatment rather than only those whose treatment resulted in triplet pregnancy. The increase in numbers could compensate for any 'weakening' of data that may result from the use of batches.

3. **Further validation of results:** The data from 150 batches of embryos was analysed using C4.5 to obtain the 'best' rule set. The method described in Chapter 3 (Section 3.3.4) provided a way of obtaining an estimate of the accuracy of this rule set on new data (i.e. by the method of cross validation). In this Chapter, the veracity of this estimate is attested further by examining the performance of the rules on new data from 50 more batches (validation set).

5.2 Organisation of this chapter

The chapter has 4 main sections, namely, Materials and Methods, Results, Discussion and Conclusions. The section on Materials and Methods (Section 5.3) details the following: (a) the 53 features selected; (b) selection of the patients; (c) collection of data from patient files; (d) conversion of data to a form ready for analysis by the program C4.5; and (e) the procedure for obtaining the rules. The composition of one particular feature, namely, the embryo grade used by the embryologists is singled out for special attention in Section 5.3.7. Section 5.4 outlines the scope of this study followed by Results that are presented in Section 5.5 and interpreted in Section 5.6. This is followed by a comparison in Section 5.7 of results obtained using decision trees with those obtained using other statistical methods to analyse the same data. The conclusions of the 'batch' study are in Section 5.8

5.3 Materials and methods

Figure 5.1 gives a 'road map' of the procedure followed in this study. Explanations of the steps involved are in the sections sign-posted in this figure.

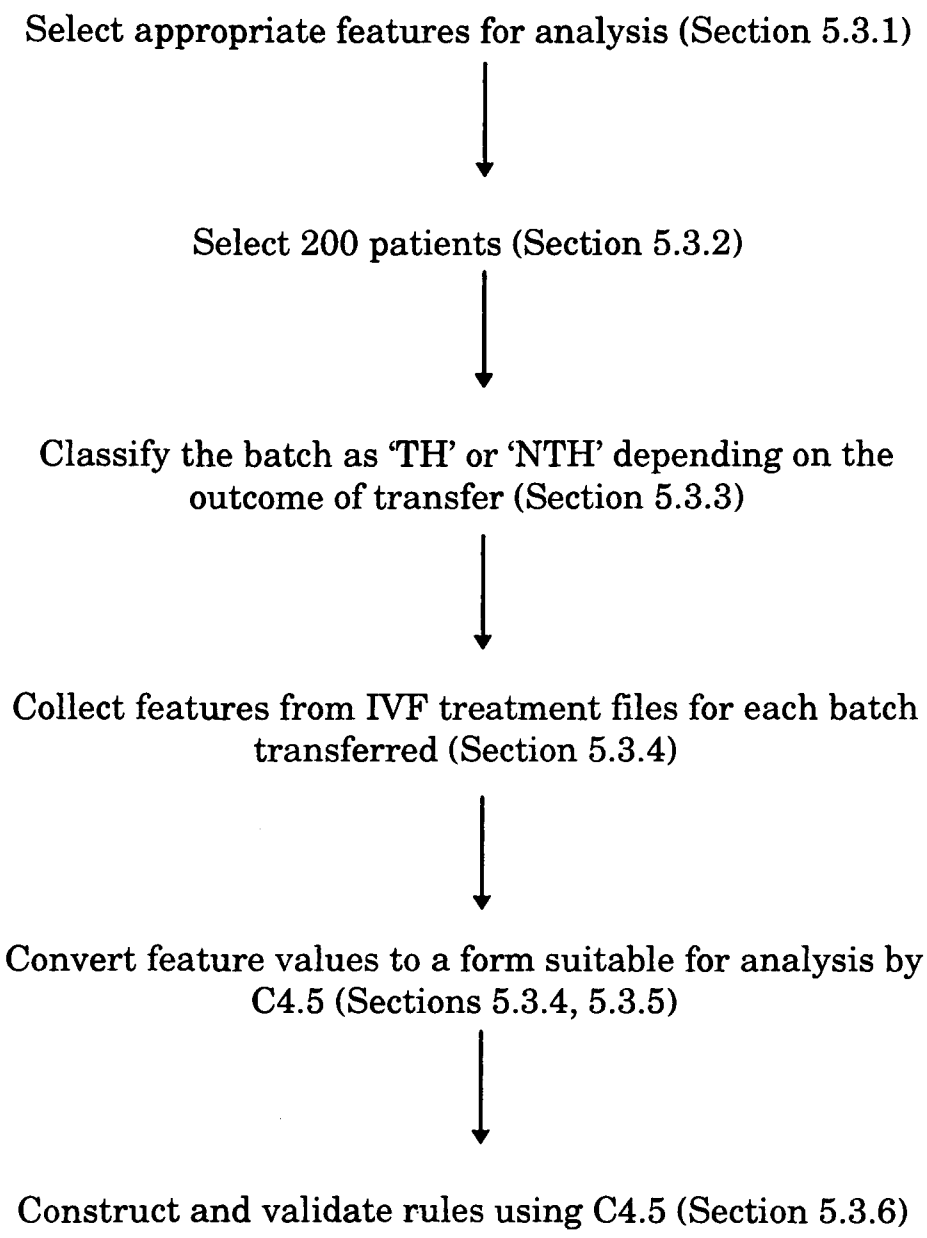


Figure 5.1 Outline of the Materials and Methods

5.3.1 Feature selection.

A number of features relating to the oocyte, follicle, sperm and embryo as well as the patient and the partner (or donor) are recorded by the IVF Unit staff in the patient's treatment record sheets. Like the pilot study,

only embryo-related (i.e. follicle, oocyte and embryo) features that could be traced to individual embryos, were chosen for inclusion in the analysis. Unlike the pilot study which concentrated mainly on features used for the selection of embryos for transfer, the batch study included features that were recorded during various stages of the IVF process irrespective of their role in the current selection process. Thus a total of 53 features (detailed in Section 3.4 and tabulated in Appendix A) were selected. Of these, 36 described the embryo, 15 the oocyte of origin and 2 described the follicle of origin of the oocyte.

5.3.2 Patient selection

IVF patients who received treatment between June 1992 and November 1993 were considered for inclusion in this study. The intention was to include eligible patients most recently treated in the IVF Unit. November 1993 was chosen as the cut-off as at the time of data collection, records for pregnancy outcome were only available for patients who had had their IVF treatment cycle on or before this date. Of the patients considered, certain pre-decided categories (see Table 5.1) were excluded and the 200 patients who were eligible were included in the study.

Criteria for exclusion of patients in this study

Patients excluded from this study fell into 8 categories as shown in Table 5.1. Categories 3-6 indicate patients who had to be excluded

because some of the 53 features chosen for inclusion could not be traced to the individual embryo.

Table 5.1 Criteria for exclusion of patients

No	Criterion	Reason
1	Less than 3 embryos transferred	to keep the sample uniform as 3 embryos were transferred in the majority of patients.
2	IVF treatment outcome being a biochemical pregnancy, spontaneous abortion or stillbirth.	the embryos in such cases may be of an intermediate quality as compared to those resulting in a 'negative' pregnancy test or a 'take home baby'.
3	Presence of more than one oocyte in the follicular fluid (ff) due to the mistaken aspiration of more than one follicle at a time.	it is not possible here to trace each oocyte to the follicle of origin. Information on the follicle is thus lost.
4	Oocytes obtained in the fluid aspirated from the Pouch of Douglas. ff could collect in the pouch as a result of follicular rupture which may occur due to trauma by the aspirating needle.	it is not possible to trace each oocyte to its follicle of origin. Information on the follicle is thus lost.
5	Oocytes obtained from 'resticks' where the aspirating needle is reinserted into follicles which originally did not yield an oocyte but subsequently re-expanded.	it is not possible to be sure of the size of the follicle as 'resticking' is based on the memory of the position of a follicle which did not yield an oocyte. A 'restick' could occasionally be the aspiration of a new follicle.
6	Insemination of multiple oocytes in a single dish as a result of a low sperm count.	it is not possible to trace each resulting embryo to its oocyte of origin thus losing information on the oocyte as well as the follicle containing the oocyte.
7	Patients lost to follow up, such that the final outcome of the IVF treatment is unknown.	self-evident
8	Patients files not traceable.	self-evident.

5.3.3 Batch selection and classification

Each batch of 3 embryos transferred to each of the 200 patients was included. Each batch was classified into one of 2 classes: 'TH' or 'NTH' depending on the outcome of IVF treatment. Here 'TH' signifies that the patient had a term or near term pregnancy, with the delivery of one, two or three live babies. 'NTH' indicates that the patient had a negative pregnancy test. For the data here, the 'TH' class had 60 and the 'NTH' class had 140 batches of embryos.

The TH rate for the 200 patients included was therefore 30%. This is higher than the average rate observed in the IVF clinic (Take Home Baby rate for 3639 Analogue cycles, 1987-94 = 22.6%). The high rate observed here could be because the sample used was selective and excluded the categories listed in Table 5.1.

5.3.4 Data collection and conversion

Data relating to the 53 selected features for each of the 600 embryos (200 batches) was collected. As mentioned in the pilot study, some features were recorded as simple 'yes' or 'no' values; others were recorded as qualitative descriptions; and finally some features were recorded as numerical values.

Unlike the pilot study, where the values for individual embryos were analysed, here each batch of embryos was to be represented by an

average value for each feature, over the 3 embryos in the batch. Calculation of the average for a feature depended on the 'scale of measurement' of that feature (namely nominal, ordinal or ratio: see Section 4.3.3). Ordinals were given numeric codes to allow analysis by C4.5 (Section 4.3.3). Following this, the average for each feature was calculated as detailed in Section 5.3.5 below.

The features used for analysis were described in detail in Chapter 3 (see Section 3.4). For completeness, a full list of the 53 features along with their levels of measurement and number or code is available in Appendix A.

5.3.5 Batch average

Depending on the scale of measurement of a feature, the mean, median or the mode was used to obtain the average value for the feature. For nominals, the mode, for ordinals, the median and for ratio measurements, the mean were calculated[1].

Table 5.2 gives one example of this calculation. This procedure gave one average value for each feature for a batch of 3 embryos. Table 5.3 shows the first few items of data in the final form ready to be used as the input for C4.5.

---

[1] If due to missing values the median or mean could not be calculated, the mode was used. If in nominals all 3 values were different (yes, no, ?), then the mode was entered as unknown, i.e. ?

Table 5.2 An example of the calculation of the batch average

	Coded value of features for each embryo				
Embryo	Follicle size in mm (ratio)	Nuclei visibility (nominal)	Cell regularity (ordinal)	Cumulus darkness (ordinal)	...
Embryo 1	16.75	yes	3	2	...
Embryo 2	18.00	no	4	1	...
Embryo 3	17.50	no	1	2	...
Batch average	17.45 (mean)	no (mode)	3 (median)	2 (median)	...

Table 5.3 Examples of data in the final form ready for analysis

	Average value of feature for the batch				
Batch	Follicle size in mm	nuclei visibility	Cell regularity	...	Class
Batch 1	17.45	no	3	...	NTH
Batch 2	14.45	no	4	...	NTH
Batch 3	18.00	yes	2	...	TH
Batch 4	16.67	yes	3	...	NTH
...	...	...	...	...	...
Batch 'n'	19.25	no	4	...	NTH

NTH = 'No take home baby' batch
TH = 'Take home baby' batch

5.3.6 Obtaining rules

The method for obtaining and validating the rules proceeds along the 4 steps shown in Figure 5.2. Explanations for each step in the sequence follow.

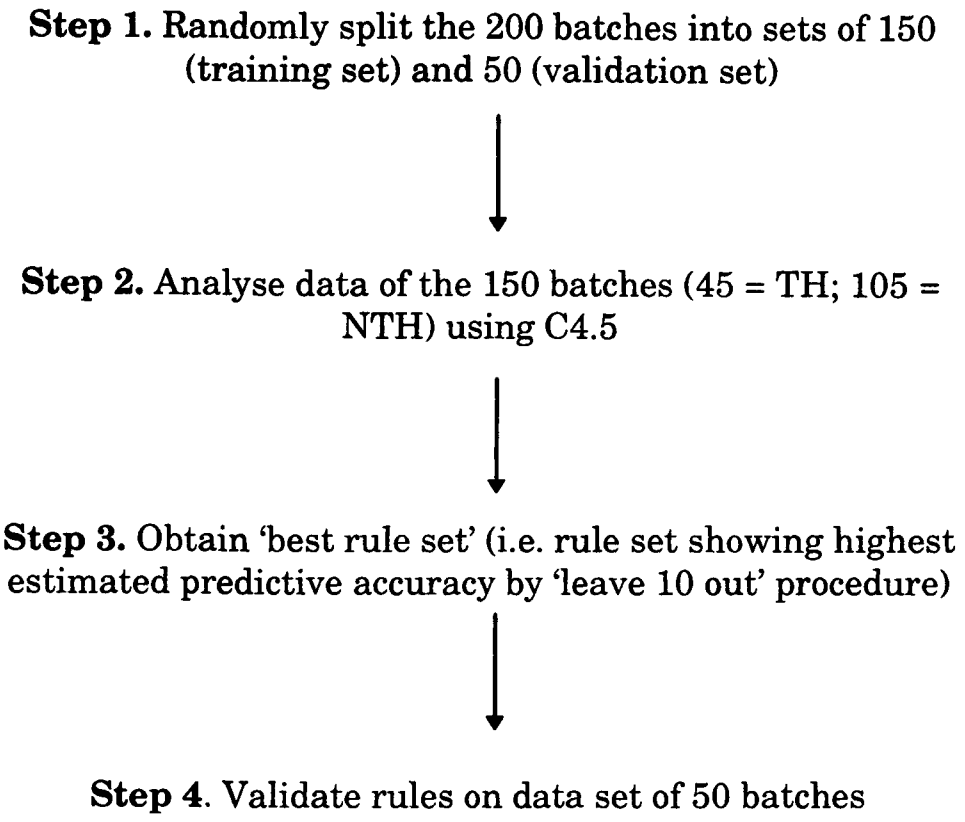


Figure 5.2 Steps involved in constructing and validating rules

Step 1. The 200 batches of embryos were randomly divided into a set of 150 batches for analysis and another set of 50 batches for validation (explained in Step 4 below). The split is achieved by a computer program[2] that randomly selects for use in Step 4, 50 of the 200 batches.

Step 2. The data for the 150 batches was used by C4.5 to construct the 'best' rule set i.e. the rule-set with the 'highest' estimated predictive accuracy (see Section 3.5.4).

Step 3. The procedure to obtain the 'best' rule set has been described in Chapter 3 (see Sections 3.5 and 3.6) and is reiterated here. A set of trees from complex to simple were constructed (using differing settings for the

[2] Written by Dr Ashwin Srinivasan of the Oxford University Computing Laboratory.

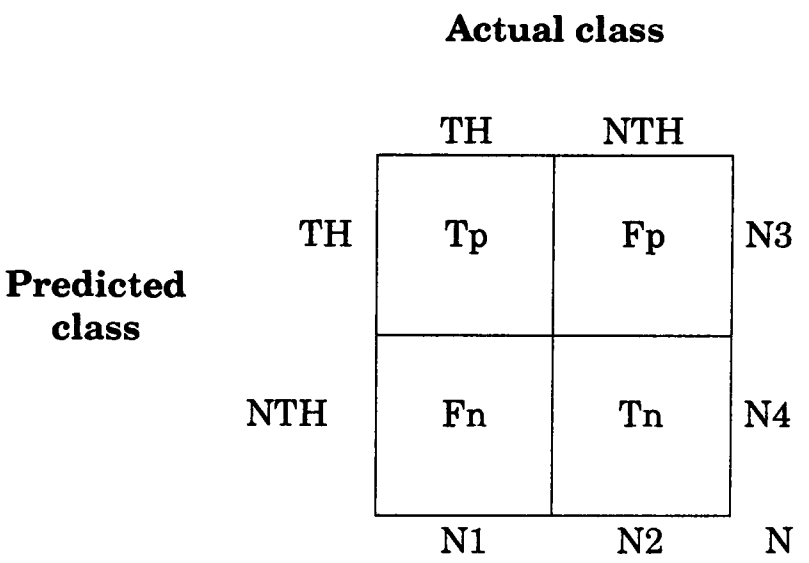
C4.5 parameters 'm' and 'c') using all the cases (150) in the sample. For each tree:

- a C4.5 rule set was obtained.
- A cross validation 2×2 table was obtained for the rule set by the 'leave 10 out' procedure.
- The χ^2 value for the table was calculated. For rule sets with a χ^2 >3.84, the estimate of predictive accuracy was recorded.
- The rule set that had the highest estimated predictive accuracy was the 'best' rule set.

Step 4. The 'best' rule set was then used to classify the cases in the validation set of 50 batches and the accuracy of this classification was calculated. This step is not an essential step and was therefore not mentioned in Chapter 3. It was carried out here to confirm that the accuracy of classification of a new data set (i.e. the predictive accuracy) would be within 2 standard deviations of the estimate of predictive accuracy (see Section 3.5.4).

An additional statistic which has not been mentioned in the methods in Chapter 3 will also be reported here. This is the 'estimated TH predictive accuracy' which gives an idea of the ability of the pattern identified to classify the TH batches accurately. This is the proportion of batches that actually resulted in a TH, of the total number of batches predicted by the pattern as TH. Consider the 2×2 table below, in which the entries have

the following meaning: Tp or true positives, Fp or false positives, Fn or false negatives, and Tn or true negatives.



Estimated TH Predictive Accuracy = Tp/N3.

This should not be confused with the estimated predictive accuracy which gives the value for the correct overall classification of all (TH and NTH) batches calculated as (Tp + Tn)/N.

5.3.7 What does 'embryo grade' capture?

Embryo grade is one of the 53 features used in the analysis. It is the grading given by the embryologist to each embryo and involves the composite assessment of 5 features (as described in Table 3.5). While each of these features has been separately included in the analysis, the importance of these features individually could be masked by the embryo grade. In order to understand better the nature of this composite assessment, the analysis by C4.5 was repeated on the data after excluding the embryo grade feature. Any feature or features that appeared in the

second set of rules in place of the embryo grade could then be construed as representing the effect of the embryo grade feature.

5.4 Scope of the study

It is important to note that all data for this study are in some sense 'filtered' by an embryologist as data used to construct the patterns, that is, the 'training data', relate only to batches of embryos selected for transfer (see Figure 5.1)

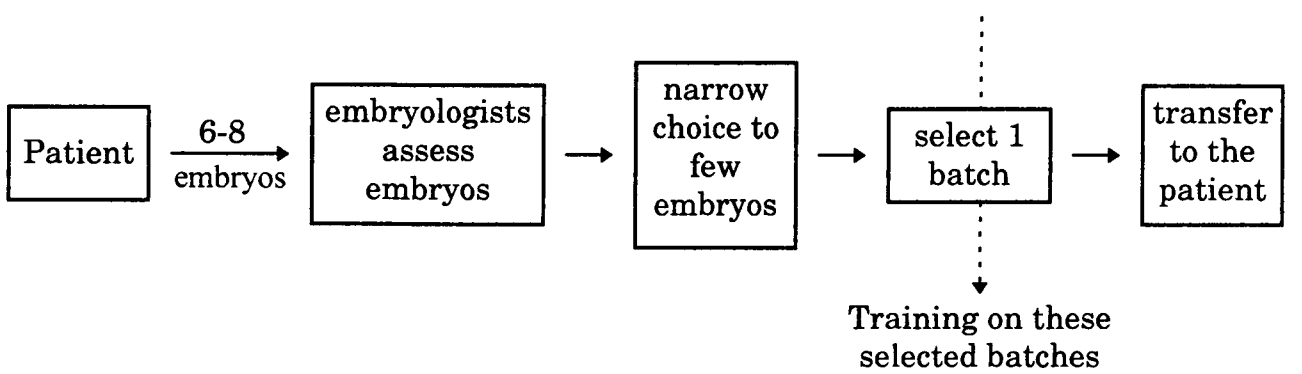


Figure 5.3 'Training data' comprised of batches selected by embryologists for transfer

As shown in the figure above, embryologists assess the available embryos and using the criteria described in Section 3.4.5 narrow their choice to a small number of embryos which they consider as having a 'good' developmental potential. Of these, three (constituting a batch) are then selected for transfer.

Based on existing knowledge of embryo development and their clinical experience, an embryologists view of embryo-related features can be grouped as in Table 5.4

Table 5.4 Categories of features pertinent to embryo selection

Category	Role of features	Features
A	Features used by embryologists to select 'good' embryos for transfer	*follicle size*: preferably 16-18mm *corona darkness*: okay, light or slightly dark *nuclei* (check 2): visible *embryo grade*: preferably grade 1-3, using the 5 component features: *cell number, regularity, fragmentation, darkness* or *granularity* all assessed at check 2 i.e. feature nos. 1,6,43,53 (40-42,44,45: components of 53) in Appendix A
B	Features and feature values used by embryologists to exclude 'bad' embryos	*follicle size*: <16 and >18 usually not preferred *corona darkness*: dark, borderline, very dark *abnormalities* of the *oocyte* or *embryo* *nuclei*: (check 2): not visible *embryo grade*: >3, using the 5 component features: *cell number, regularity, fragmentation, darkness* or *granularity* all assessed at check 2 i.e. feature nos. 1,6-15,17-25,31-39,43-53 (40-42,44,45 components of 53) in Appendix A
C	Features that are normally not considered when selecting embryos for transfer	*ff volume, oocyte cumulus amount and description, oocyte corona maturity, cumulus dish attachment; cell number, regularity, fragmentation* and *nuclei* at check 1 i.e. feature nos. 2-5,16,26,27,28,29,30 in Appendix A

In an ideal setting, we can imagine that embryos with features in category B will not be selected for transfer by an embryologist. Instead choice will be guided largely by features in category A. Occasionally in situations where sufficient embryos with these values are not available, other embryos may be selected. Since embryologists however, try to restrict the transferred embryos to those satisfying category A features, and all

embryos used here have been selected by embryologists, we would expect very few batches to contain embryos with category B features.

How could a study such as this, contribute to existing knowledge as embodied in Table 5.4? It would be unreasonable to expect that the results will have much to say concerning category B features and their relationship to developmental potential - only because of poor representation in the data. The results could make the following contributions:

- Identify if all category A features are truly relevant.
- Refine and clarify the values of relevant features for 'good' embryos in category A, e.g. follicle size between 16-18mm is thought to be important. If it is, can one be more precise?
- Investigate if all category C features are truly irrelevant. If not, identify the relevant features and their relationship to development potential.
- The 'best rule set' obtained could be used to assist embryologists in the selection of batches for transfer. For the estimates of predictive accuracy and TH predictive accuracy obtained to be reliable, the rules must be applied to data that are broadly similar to that used for training, namely batches that embryologists deemed worthy of transfer. That is, the rules could be used to select a batch from several whose TH potential would be indistinguishable for an embryologist as shown in Figure 5.4. It is conceivable that any general principles identified by a

set of rules could be used to guide the selection of individual embryos. This is considered further in a later section.

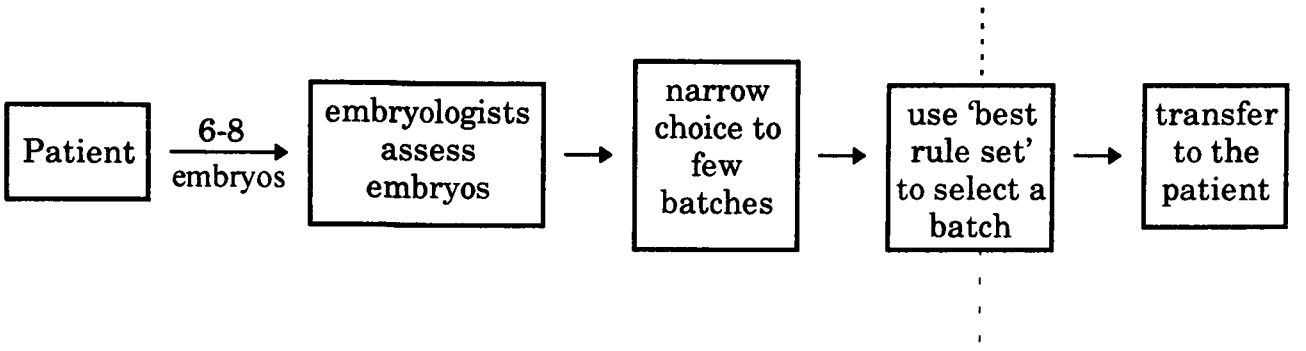


Figure 5.4 Stage in the selection procedure at which rule set may be used

The rule set could therefore assist selection only in that 'grey area' where embryologists have narrowed their choice, but of these embryos are not sure as to the batch that has the best chance of resulting in a pregnancy. In some cases, the rules might not identify any of the given batches as having the characteristics of a TH batch. In such a situation, the batch that the embryologists would have normally transferred could still be transferred.

5.5 Results

Results are presented as follows. First, the rules obtained are shown. Next the 2×2 table representing the performance from 'leave 10 out' cross validation, its χ^2 value, estimate of predictive accuracy with its standard deviation and an estimate of the TH predictive accuracy are given. This is

followed by a 2×2 table representing the performance on the validation set of 50 batches.

Following the above protocol, Rule Set 1 obtained by the analysis of all 53 features is presented first. Rule Set 2 obtained by excluding the feature embryo grade from the analysis, is presented next. It must be remembered that the analysis has been conducted on the average values calculated for every batch of 3 embryos.

5.5.1 Results obtained using all 53 features

Rules

Rule 1

If
  embryo grade = 1 and
  follicle size > 16.33 mm and
  follicular fluid volume > 2.75 ml and
  follicular fluid volume ≤ 4.25 ml
Then
  Class = **TH**

Rule 2 (default rule)

If
  none of the above
Then
  Class = **NTH**

Henceforth, this set of rules will be referred to as Rule Set 1.

Table A

'leave 10 out' cross validation 2×2 table

		Actual class: TH	Actual class: NTH	
Predicted class	TH	17	11	28
	NTH	28	94	122
		45	105	150

χ^2 value = 13.7 (P<0.001)

Estimate of predictive accuracy of the rule set = 74.0% (± 3.6)

Estimate of TH predictive accuracy = 62.1% (± 9.2)

Note that Rule Set 1 corresponds essentially to the existing embryologists' selection criteria with an important refinement requiring just two of the remaining 52 features. The results shown in Table A raise the possibility that if embryo batches were selected for transfer using this refinement i.e. embryo grade = 1 together with follicle size and volume, then substantially improved take-home baby rates (62.1%) may be expected in that sub-group of patients who had a batch satisfying these conditions. This expectation is subject to sampling error and to the qualifying presumption that new cases of broadly similar characteristics are treated under similar conditions. The testimony of Table B which summarises a wholly independent subset of the data, gives further support to this.

Table B

2×2 table obtained for the validation set of 50 batches held aside from the initial 200.

		Actual class		
		TH	NTH	
Predicted class	TH	6	4	10
	NTH	9	31	40
		15	35	50

Predictive accuracy = 74 %

TH predictive accuracy = 60%

5.5.2 Results obtained excluding embryo grade

Rules

Rule 1

If
cell number (check 1, day 2) > 3 and
follicle size > 16.33 mm and
follicular fluid volume > 2.75 ml and
follicular fluid volume ≤ 4.25 ml
Then
Class = **TH**

Rule 2 (default rule)

If
none of the above
Then
Class = **NTH**

Henceforth, this set of rules will be referred to as Rule Set 2.

Table A

'leave 10 out' cross validation 2×2 table

		Actual class: TH	Actual class: NTH	
Predicted class	TH	22	11	33
	NTH	23	94	117
		45	105	150

$\chi 2$ value = 24.9 ($P<0.001$)

Estimate of predictive accuracy = 77.3% (± 3.4)

Estimate of TH predictive accuracy = 66.7% (± 8.2)

The implication of the results obtained with Rule Set 2 is that the important aspect captured by the embryologists' grading seems to pertain to the rate of cleavage or the cell number. Replacing embryo grade = 1 by cell number > 3 is not shown to cause any significant alteration in either the estimated predictive accuracy or estimated TH predictive accuracy rates. The testimony of Table B gives further support to this.

Table B

2×2 table obtained for the validation set

		Actual class: TH	Actual class: NTH	
Predicted class	TH	6	4	10
	NTH	9	31	40
		15	35	50

predictive accuracy = 74 %

TH predictive accuracy = 60%

5.5.3 Further analysis

This section includes two analyses conducted on the rule sets presented in Sections 5.5.1 and 5.5.2 above. The first investigates the importance of the ff volume component of the rule sets. The second investigates the consequences of relaxing the precise cut-off values for follicle size and ff volume in the rules by rounding them to whole numbers.

Assessment of ff volume component of rule sets

An investigation was conducted to confirm the importance of the feature ff volume that was identified in both rule sets, for the following reasons:

1. **Correlation with follicle size**: Follicle size would be expected to bear a strong correlation with the ff volume. In the presence of follicle size, the feature ff volume may not therefore contribute towards any substantial increase in the estimated predictive accuracy.

2. **Feature picked up last:** Since the ff volume is the third feature identified in the rules, this indicates that it was used to split the classes in one of the nodes lower down in the decision tree. The number of cases in each node decreases the further down the node is. With small numbers, a feature could be erroneously identified as being important. It is desirable therefore to investigate its importance by an independent analysis as shown below.

Given the relationship between follicle size and ff volume, it may be expected that the lower bound (i.e. follicular fluid volume > 2.75) would be sufficiently captured by the follicle size >16.33 component. The ability of the rule sets to classify batches, when the lower bound for ff volume was excluded, was investigated first. This was followed by an investigation into the ability of the rule sets to classify batches when the ff volume component was completely excluded. The predictive accuracy and TH predictive accuracy values were estimated by using the modified rule sets to classify new data, in this case, the 50 batches in the validation set[3]. The results are presented in Table 5.5 and Table 5.6.

---

[3]It is not possible to obtain such estimates from the 150 batches as the modifications do not represent the rules obtained from any particular setting for C4.5.

Table 5.5 Assessment of importance of ff volume in Rule Set 1

Modification	Predictive Accuracy (%)	TH Predictive Accuracy (%)
Complete rule set embryo grade = 1 and follicle size > 16.33 mm and follicular fluid volume > 2.75 ml and follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	74	60
Excluding lower bound of ff volume embryo grade = 1 and follicle size > 16.33 mm and follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	74	60
Excluding ff volume entirely embryo grade = 1 and follicle size > 16.33 mm → TH none of the above → NTH	74	57

The results in Table 5.5 and Table 5.6 should be interpreted cautiously, given the relatively small number of batches involved. With this in mind, the tabulations indicate that the exclusion of the lower bound of the ff volume (>2.75ml) makes no difference to the accuracies and it can therefore be excluded from the rule sets. Exclusion of the upper bound (≤ 4.25ml) however, causes a small fall in the predictive accuracy of Rule Set 2 (from 74 to 72) as well the TH predictive accuracy of both rule sets (from 60 to 57 for Rule Set 1 and from 60 to 53 for Rule Set 2). Whilst the fall appears to be relatively insignificant here, caution dictates that the upper bound on ff volume be retained until the effect of its removal is shown to be minimal on a larger data set.

Table 5.6 Assessment of importance of ff volume in Rule Set 2

Modification	Predictive Accuracy (%)	TH Predictive Accuracy (%)
Complete rule set cell number (check 1, day 2) > 3 and follicle size > 16.33 mm and follicular fluid volume > 2.75 ml and follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	74	60
Excluding lower bound of ff volume cell number (check 1, day 2) > 3 and follicle size > 16.33 mm and follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	74	60
Excluding ff volume entirely cell number (check 1, day 2) > 3 and follicle size > 16.33 mm → TH none of the above → NTH	72	53

Are the precise cut-off values for follicle size and ff volume important?

The second analysis reported in this section was conducted to investigate if altering the values of the follicle size and ff volume identified here made any difference to the predictive accuracy and TH predictive accuracy. As with the previous analysis, this was done by testing the ability of the modified rule sets to classify new data (i.e. the validation set of 50 batches), and the same warning regarding small data sizes applies. In light of the results obtained above, the lower bound for ff volume has been excluded from the analysis here. The results are presented in Table 5.7 and Table 5.8.

Table 5.7 Relaxing follicle size and ff volume values in Rule Set 1

Modification	Predictive Accuracy (%)	TH Predictive Accuracy (%)
Rule set with precise values embryo grade = 1 and follicle size > 16.33 mm → TH follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	74	60
Relaxed follicle size value embryo grade = 1 and follicle size > 16 mm → TH follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	66	42.9
Relaxed follicle size value embryo grade = 1 and follicle size > 17 mm → TH follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	66	33.3
Relaxed ff volume value embryo grade = 1 and follicle size > 16.33 mm and follicular fluid volume ≤ 4 ml → TH none of the above → NTH	72	55.5
Relaxed ff volume value embryo grade = 1 and follicle size > 16.33 mm → TH follicular fluid volume ≤ 5 ml → TH none of the above → NTH	72	54.5

Although a relaxation in the follicle size cut-off from 16.33mm to 16.4mm causes no alteration in either the predictive accuracy or the TH predictive accuracy values,. the results presented in Tables 5.7 and 5.8 indicate that a relaxation in values for follicle size to 16 or 17mm could result in a substantial lowering of both accuracies. It must be remembered however,

Table 5.8 Relaxing follicle size and ff volume values in Rule Set 2

Modification	Predictive Accuracy (%)	TH Predictive Accuracy (%)
Rule set with precise values cell number (check 1, day 2) > 3 and follicle size > 16.33 mm → TH follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	74	60
Relaxed follicle size value cell number (check 1, day 2) > 3 and follicle size > 16 mm → TH follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	66	42.9
Relaxed follicle size value cell number (check 1, day 2) > 3 and follicle size > 17 mm → TH follicular fluid volume ≤ 4.25 ml → TH none of the above → NTH	66	33.3
Relaxed ff volume value cell number (check 1, day 2) > 3 and follicle size > 16.33 mm **and** follicular fluid volume ≤ **4** ml → TH none of the above → NTH	74	62.5
Relaxed ff volume value cell number (check 1, day 2) > 3 and follicle size > 16.33 mm → TH follicular fluid volume ≤ **5** ml → TH none of the above → NTH	72	54.5

that in clinical practice precise measurement to within 1mm of follicle size may not be possible given constraints of speed and currently available methods. The importance of these results therefore lies chiefly in suggesting that a progressive lowering of accuracy could be expected to occur as individual embryos in a batch are associated with progressively

larger or smaller follicles. This could be demonstrated by investigating the effect of progressivley lowering the average follicle size to 15, 14, 13mm and lesser or progressively increasing average follicle size to 18, 19, 20mm and beyond in the rule sets. The size of the validation set here is however too small (50 matches) to demonstrate clearly associated changes in accuracy. In future this could be investigated using a larger size data set.

The precise value of the threshold on ff volume, at 4.25ml, appears to be less critical, with relaxations to 4ml or 5ml causing what appear to be insignificant changes to both accuracies. On the basis of the tabulation here, a ff volume threshold of 4ml appears to be just as acceptable as the more precise threshold identified automatically. As with follicle size however, progressively increased or lowered values would be expected to cause a progressive lowering of accuracy.

5.6 Discussion

The relationships of 53 features related to the embryo, oocyte and follicle of origin to the developmental potential of the embryo were investigated using C4.5. The analysis yielded 2 rule sets that identified the patterns of features characterising embryos with a TH or a NTH potential.The results obtained are examined considering each rule set in turn.

5.6.1 Rules obtained using all features (Rule Set 1)

Rule Set 1 obtained using all 53 features, identifies 3 features namely, embryo grade, follicle size and ff volume as bearing an important

relationship to development. The identification of embryo grade as an important feature provides quantitative verification of the embryologists' intuition. Batches of embryos with an average embryo grade of 1 were shown to have a higher potential of resulting in a TH baby. Embryo grades are however assigned by the embryologist after considering 5 features, namely, cell number, fragments, regularity, darkness and granularity of the cytoplasm just prior to transfer (that is, check 2 on day 2). With reference to the embryologists grading scheme (see Table 3.5) it is evident that grade 1 embryos are characterised by 4-5 cells (i.e. a rate of cleavage that is neither too slow, nor too fast), that are fairly regular and have few or no fragments in the perivitelline space. It is not immediately obvious from Rule Set 1, that all these features have an equally important relationship with the developmental potential. The discussion in the following section (Section 5.6.2), of Rule Set 2, which was constructed in order to identify the important features captured by embryo grade, is pertinent to this issue.

The identification of follicle size and ff volume indicates that the oocyte stage plays an important role in determining the developmental potential of the embryo. A number of studies have reported that follicle size and ff volume reflect the maturity of the oocyte of origin of the embryos (Scott *et al*, 1989; Simonetti *et al*, 1985). With the increase in ovarian stimulation, the size (and volume) of the follicles gradually increases reflecting the progressive maturation of the contained oocyte. The rule set obtained here, identifies an average follicle size > 16.4mm as one of the key

determinants of the potential of a batch of embryos to result in a TH outcome. During the selection procedure, embryologists at the Oxford Unit consider a follicle size in the range of 16-18mm as a determinant of the ability of the embryo to progress to a clinical pregnancy. The rules obtained here provide a precise value for this feature, the precise nature having been demonstrated by the drop in accuracy that occurred when the cut-off value for follicle size was rounded to a whole number. Given the data at hand therefore, 16.4 mm reflects the follicle size that indicates a level of maturation of the contained oocyte compatible with a successful pregnancy.

Other studies support the finding that embryos from very small oocytes have a poor developmental potential. Lopata (1982) showed that only oocytes from larger size follicles resulted in pregnancies. Although immature oocytes (Prophase I in Veeck, 1988) are capable of undergoing fertilisation and cleavage, the resulting embryos have been shown to have no potential (or a very low potential) for developing into a pregnancy (Zenzes *et al*, 1985 and Veeck, 1988). Zenzes and colleagues suggest that one possible reason for this could be the marked asynchrony in pronuclear morphogenesis causing developmental arrest. Simonetti *et al* (1985) also found that the rate of abortion was very high when the conceptus was derived from follicles of a smaller size (< 2.0 ml ff aspirated)

An average ff volume between 2.75ml and 4.25ml was also identified in the original rule set. Further analysis however showed that the lower

bound of ff volume was redundant in the presence of follicle size, given the expected correlation between the two features (assuming a spherical shape for the follicle, size of 16.4mm is equivalent to around 2.3ml ff volume). The upper bound of ff volume of ≤4.25ml however, complemented the follicle size. Further, this could be relaxed to 4ml without a significant lowering in the predictive accuracy. While the follicular size cut-off value of >16.4mm therefore excludes immature oocytes, the ff volume value of 4ml excludes post-mature oocytes from the high potential group. Nayudu *et al* (1987) also showed that the majority of normal pregnancies resulted from follicles that had an intermediate volume, i.e. those containing mature rather than immature or post-mature oocytes. Veeck (1988) showed that oocytes of more advanced maturity resulted in greater pregnancy loss after implantation. Follicles with a higher ff volume are likely to have been stimulated for a longer duration resulting in aged post-mature oocytes - with a higher likelihood of defects of the meiotic apparatus (Jongbloet and Zwets, 1976).

Embryologists do not consider ff volume when selecting embryos and recognition by the decision tree technique of an important role for this feature, indicates that it is worthy of further investigation. The fact that ff volume is identified after follicle size in the rule set and that the lower bound could be dropped without a reduction in accuracies, however, gives an indication that follicle size was considered a more predictive feature than ff volume. This does actually reflect the clinical situation where

recorded measurements of follicle size are considered more reliable than ff volume. This is because firstly, all the fluid in the follicle may not have been aspirated. Secondly, measurements of volume of fluid aspirated are not taken very precisely due to the need for speed in moving the recovered oocyte into the 37ºC incubator.

5.6.2 What does 'embryo grade' capture (Rule Set 2)?

Rule Set 2, namely that obtained by excluding embryo grade from the analysis, also identifies 3 features as bearing an important relationship to developmental potential. These are: cell number assessed at the first check on the day of transfer, follicle size and ff volume.

On comparison of Rule Sets 1 and 2, the identification of cell number at check 1, in place of embryo grade, indicates that while embryologists arrive at a value for embryo grade using a (relatively) complex assessment of 5 other embryo features, the main aspect captured pertains to the cell number i.e. the rate of cleavage. This is further supported by the finding that Rule Set 2, has exactly the same accuracy of classifying embryos in the validation set (74%) as Rule Set 1.

Rule Set 2 indicates that batches with an average cell number > 3 at check 1 would have a higher potential of resulting in take home babies. A number of studies have similarly reported on a higher IVF success rate when embryos with a faster rate of cleavage were chosen for transfer (Staessen *et al*, 1992; Claman *et al*, 1987; Edwards *et al*, 1984; Trounson,

1983 and Mohr *et al*, 1983). Two studies however, suggest that too rapid a rate of cleavage could actually prove detrimental (Cummins *et al*, 1986; Testart, 1986). The rule set here does not actually define an upper limit to the cell number. It is highly probable however that when the rule cell number > 3 is used in conjunction with the other rules namely follicle size < 16.33mm and ff volume > 4.25ml, these would ensure that embryos with a very rapid cleavage rate would be excluded from the batches of embryos identified as having a good developmental potential.

The importance of follicle size and the upper bound for ff volume remain undiminished in Rule Set 2 with the same values as those identified in Rule Set 1.

The identification of a number of features in the rule sets obtained suggests that the interaction between features is of key importance. This indicates that developmental potential is not determined by a single feature but is influenced other features. All embryos originating from oocytes of similar maturity status therefore do not essentially have a similar developmental potential. The oocytes could be influenced by the *in vitro* culture conditions and sperm used for insemination resulting in embryos of differing appearances and rates of cleavage as well as a differing developmental potential.

Most studies to date with the notable exception of Nayudu *et al* (1987 and 1989) have studied either the correlation between oocyte stage and pregnancy outcome or embryo stage and pregnancy outcome. The present

study however shows how such an approach ignores the influence of each stage on the other in finally determining the pregnancy outcome.

5.6.3 Relevance to current assessment procedures

The experimental design of this study (i.e. the use of data for transferred batches of embryos) allows for the comparison of the results with the features considered important by the embryologists when selecting embryos for transfer. With respect to this, and with reference to Table 5.4 the following comments can be made:

1. The results of the batch study give an objective and unbiased corroboration of the embryo assessment methods used in the Oxford IVF Unit. Embryo grade and follicle size are considered two of the most important features when selecting bipronucleate embryos for transfer. Of the category A features in Table 5.4, this study shows that the features embryo grade and follicle size together have a strong correlation with the developmental potential of embryos. This therefore supports strongly the assessment of these features when selecting embryos for transfer.

2. Currently, embryos with a grade from 1-3 (grade 1 is considered the best) and follicle size between 16-18mm are preferentially selected over others. The batch study refines and makes quantitatively precise the values of these features.

3. In addition to embryo grade and follicle size, selection involves an assessment of the darkness of the corona and visibility of nuclei in the cells of the embryo (category A features in Table 5.4). These features have however not been identified in the study as playing an important role in determining the developmental potential. This is therefore a pointer towards their relative unimportance .

4. Of the category C features in Table 5.4, the study identifies ff volume which is currently not directly considered when selecting embryos for transfer. This feature is recognised as an important determinant of developmental potential and therefore worthy of further investigation. The study also identifies cell number at the first check on the day of transfer (i.e. the rate of cleavage) as capturing substantially the information provided by the embryo grade 1.

5. This study confirms the relative unimportance of other features in category C not taken into consideration when selecting embryos for transfer.

It is interesting to abstract the general principles obtained from batch averages, into a method for guiding the selection of individual embryos. Based on the modified Rule Set 1 or 2 (i.e. after exclusion of the lower bound for the ff volume), Figure 5.5 shows a method that grades embryos from 'bad' to 'very good'.

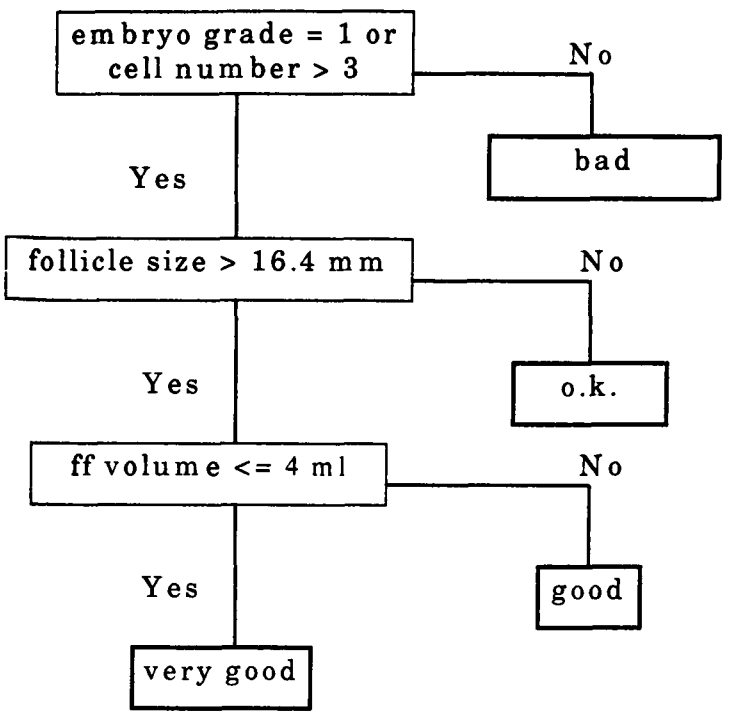


Figure 5.5 A method of grading individual embryos

The procedure outlined above can be used to select individual embryos for transfer. If individual embryos satisfy the values, the average values for all embryos in the batch would automatically fulfil the average values identified in the Rule Sets 1 or 2 for batches of embryos. Both sets of rules could also be used to select batches of embryos for transfer. As cautioned in Section 5.4, the rules will perform well only if applied to new data broadly similar to that used for training and obtaining the rules from.

The overall Take Home baby rate that these rules will result in can only be calculated after conducting a clinical trial. If conducted properly, i.e. if the application of rule sets to select embryos is restricted to a sample of patients drawn by using the exclusion criteria in Table 5.1 and the rules are used to select batches (or embryos, as suggested above) as shown in Figure 5.4, the results reported here suggest that a substantial increase in the take-home baby rate could result.

The batch study presented here however has some shortcomings, which need to be addressed before any results can be considered for clinical use.

5.6.4 Shortcomings of the experimental design of the batch study

The experimental design of this study has four main shortcomings. Firstly, the data used here comprises batches of embryos pre-selected by embryologists for transfer. Embryos with feature values considered by embryologists to be associated with a poor developmental potential, are thus not well represented in the data. The results obtained here cannot therefore clarify the relationship of these features and feature values (category B features in Table 5.4) to development. A thorough investigation of all feature values is possible only if the study is restricted to patients who have only 3 embryos available, i.e. the question of selecting the 'best' embryos for transfer does not arise. In such cases, all three available embryos are transferred irrespective of the feature values. The number of patients satisfying such a criterion was however too small to allow for a statistically valid investigation necessitating the use of the current experimental design.

Secondly, it must be remembered that in the experimental design used here, we can only be sure that the batch of embryos that resulted in a TH outcome had a good developmental potential. The batch of embryos which resulted in a negative pregnancy test however, may not inherently have a poor developmental potential. Maternal factors may have been responsible for the failure. The input to the TH pattern is therefore

'stronger' for every TH batch of embryos (with a definite good developmental potential) than that of a NTH batch (not necessarily of a poor developmental potential) to a NTH pattern. To a certain extent using the average values and a sufficient number of cases accounts for this. It would however be more appropriate if the study incorporated the issue of 'costs'. When constructing the cross-validation table therefore, the cost of a rule set misclassifying a TH batch would be considered higher than that of misclassifying a NTH batch. The costs of misclassification would then have to be taken into account when selecting the best rule set. Fixing the cost of misclassifying a TH in relation to a NTH is not straightforward issue. As an example, the cost of misclassification from a biological point of view might be completely different from that taken from the point of view of implications for IVF treatment. The costs have to be decided after consultation with embryologists and could form part of a future study which addresses the shortcomings of the study presented here

Thirdly, while the results obtained here do provide some insight into the features influencing the development of the embryo, all the features that may affect development have not been taken into consideration. In particular features related to sperm used for insemination and maternal features (other than oocyte and follicle related features) have not been considered. Embryos that have similar characteristics could differ in their developmental potential depending on the quality of the sperm sample used for insemination. Similarly, as mentioned above, although the features associated with an embryo may be suggestive of a very good

developmental potential, a hostile maternal environment could prevent it from progressing to a pregnancy. The age of the mother, the cause of infertility, the condition of the uterus, the drugs used, levels of hormones all affect the receptivity of the uterus to the implanting embryo and thus influence the embryo. The current study restricts itself to embryo-related features traceable to individual embryos and does not take into account the influence of maternal and sperm features. In the future, however, these features could be accommodated within a new experimental design and their influence on development investigated using the technique of decision trees.

Finally, the size of the data with only 200 patients (600 embryos) included in the batch study is still small by statistical standards. Increasing the data sample size will serve to refine the rules obtained, and give them greater statistical validity.

5.7 A note on using logistic regression

This section presents the results that were obtained by analysing the same data used in the earlier sections, by the logistic regression technique. The aim is to allow the reader to contrast such results against those obtained by decision trees, so that some essential differences between the two techniques in analysing such IVF data are easily apparent. The regression analysis was conducted with all 53 features in place. Standard regression practice sees the use of all of the training data to construct a mathematical equation (as described in SPSS for Windows

Advanced Statistics Release 6.0, 1993). The predictive accuracy of the equation is then estimated from new data. This method is followed here, with the 150 batches used to construct the equation and the 50 batches to estimate its predictive accuracy.

Data for the 150 training batches was analysed by stepwise logistic regression performed using the SPSS package (details of the procedure can be found in the SPSS for Windows, Advanced Statistics, Release 6.0, 1993). The equation obtained at default settings for SPSS parameters is presented below:

Regression Equation

$$\mathbf{Z} = -1.8392 \times \text{cytoplasm darkness, check 1} + 0.7232 \times \text{embryo grade} - 0.3107 \times \text{follicle size} + 5.0250$$

Usually, logistic regression analysis proceeds as follows. The **Z** value above is transformed to a **P** value that is between 0 and 1:

$$\mathbf{P} = 1/(1+e^{-\mathbf{Z}}),$$ where e is a constant and is approximately 2.7

Based on the distribution of **P** values in the training data, a threshold for discriminating between the two classes is identified. All batches with a **P** value below the chosen threshold are classified as one of the classes, and the rest are assigned to the other class.

Figure 5.6 is a histogram of **P** values of TH and NTH batches in the training set of 150 batches. The default threshold on **P** is 0.5. All batches below this value will then be classified TH, and the rest NTH. It is evident

that such a threshold for the data here would result in an unacceptably large number of TH batches being misclassified as NTH (38 of the 45 TH will be misclassified). In such situations, the method suggested is to examine alternative thresholds that result in an acceptable number of errors in one or both classes. The histograms in Figure 5.6 suggest that 0.6 is a better threshold. The classification of batches in the training set using this threshold is tabulated below.

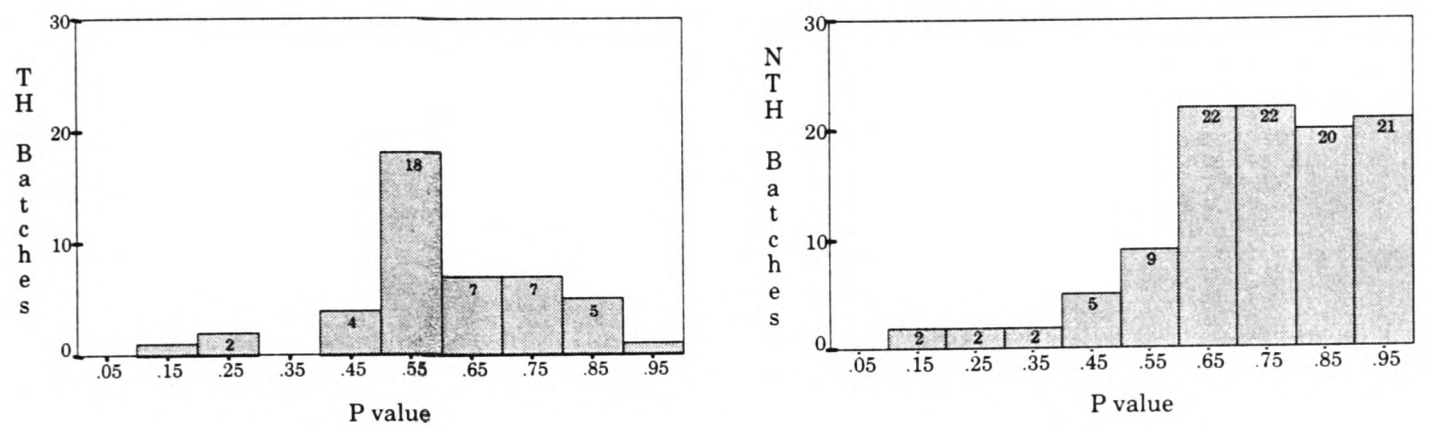


Figure 5.6 Distribution of training set batches on the basis of the P value of the regression equation

Classification of batches in the training set using a threshold of P = 0.6

		Actual class: TH	Actual class: NTH	
Predicted class	TH	25	20	45
	NTH	20	85	105
		45	105	150

Training set accuracy of the equation = 73.3% (± 3.6)

TH accuracy = 55.5% (± 7.4)

Figure 5.7 shows the histogram of **P** values of TH and NTH batches in the test set of 50 batches. Applying the same threshold (**P** =0.6), results in the classification are tabulated below. As expected, both overall and TH accuracies are lower than that obtained on the training set, and the accuracies estimated from batches in the test set suggests that for this choice of thresholds at least, the regression technique appears to perform worse than its decision tree counterpart.

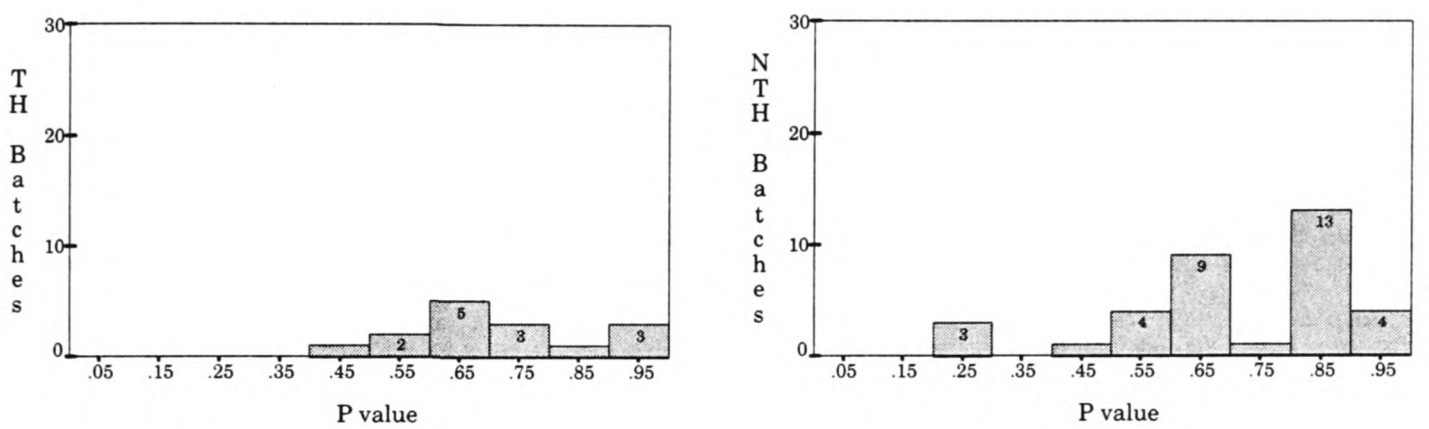


Figure 5.7 Distribution of test set batches on the basis of the P value of the regression equation

Classification of batches in the validation set using the P = 0.6 threshold

		Actual class		
		TH	NTH	
Predicted class	TH	3	8	11
	NTH	12	27	39
		15	35	50

Estimated predictive accuracy = 60.0 %

Estimated TH predictive accuracy = 27.3%

5.7.1 Relevance to results obtained with decision trees

As mentioned earlier, results here have been obtained at default settings for SPSS parameters. To allow for a proper comparison of the accuracy of these results with those from the decision tree analysis, the estimated predictive accuracy resulting from various combinations of parameter settings and thresholds would have to be examined. Although SPSS does not have an inbuilt procedure capable of doing this automatically, the default settings are decided based on good statistical reasons. Thus, while caution should be exercised when attempting comparisons of accuracy, the results can be taken to be reasonably indicative of the performance of the method of regression.

The purpose of this exercise in using logistic regression has been only to indicate the nature of the results and their difference to those obtainable by using decision tree techniques. With respect to this, the following observations can be made. Firstly, the mathematical equations bring into relief the understandability of the decision tree results. Secondly, at the heart of the method is the straight line linear equation for calculating **Z** values. The increase or decrease in the values for a particular feature can only be associated in a linear fashion, with an improvement or decline in the development potential. The rules from decision trees however delineate precise range of values for features. Combinations of these ranges are used to capture the interaction between the features and their relationship to developmental potential. It has already been mentioned

that regression techniques are poor at capturing inter-feature interactions.

5.8 Conclusion

The results obtained in this chapter by the decision tree analysis of data, identify four features (viz. embryo grade, cell number, follicle size and ff volume) as bearing an important relationship with the developmental potential of an embryo. They demonstrate how this potential is determined as early as the stage of the unfertilised oocyte. The analysis also provides evidence showing that the developmental potential is influenced by the interaction of the above-mentioned features and is not determined by individual features in isolation.

Further, the results provide corroboration of currently used methods for selecting embryos for transfer while at the same time identifying more refined and precise values of the features. Features that can safely be ignored when assessing eligibility for transfer as well as some additional features, worthy of further investigation have also been identified.

5.9 Summary

The decision tree technique was used in this chapter as a statistical tool for the analysis of data relating to batches of embryos transferred to IVF patients. The data comprised 53 features for each embryo and included features related to the embryo, as well as the oocyte and follicle of origin. The batches were classified based on the outcome of transfer into TH or

NTH. Relationships have been clarified between a number of embryo-related features and the developmental potential of a batch of IVF embryos on transfer. The results identified four important features besides offering additional insights to the developmental potential. The next chapter analyses the relationship between morphological features at the blastocyst stage and the developmental potential *in vitro* to demonstrate the applicability of the decision tree technique to other areas of IVF.

Chapter Six. Blastocyst Study

6.1 INTRODUCTION 144

6.2 ORGANISATION OF THIS CHAPTER 144

6.3 BLASTOCYST GRADING 145

6.3.1 Blastocyst Grading Scheme Of Dokras et al (1993) 147

6.3.2 Drawbacks Of The Dokras Grading Scheme 148

6.3.3 Alternate Grading Scheme Using Decision Trees 150

6.4 MATERIALS AND METHODS 152

6.4.1 Blastocyst Selection. 152

6.4.2 Feature Selection, Collection And Conversion 154

6.4.3 Obtaining Rules 156

6.5 RESULTS 157

6.6 DISCUSSION AND CONCLUSIONS 158

6.7 SUMMARY 161

6. BLASTOCYST STUDY

6.1 Introduction

The intention of this chapter is to show how the ideas developed so far can be used in other areas of IVF i.e. not just to assess the relationships of embryo-related features with development. As an illustration, the Chapter takes up the potential application of these ideas to a later stage of embryo development *viz.* the blastocyst stage. It investigates the use of these ideas for the grading of blastocysts of differing quality.

6.2 Organisation of this Chapter

The *in vitro* development of the embryo to the blastocyst stage is first briefly outlined. The need for a blastocyst grading system is then presented, followed by a discussion of the existing grading system used in the Oxford IVF Unit and its shortcomings (Sections 6.3.1 and 6.3.2). The use of decision tree analysis techniques to formulate a new grading scheme that overcomes some of these limitations is then proposed (Section 6.3.3). Section 6.4 details the selection of blastocysts of differing quality (as assessed by levels of hCG secretion) and blastocyst features to be analysed for inclusion in the new grading scheme. Section 6.5 gives the results in the form of rules defining blastocysts of differing quality. Results are discussed and conclusions arrived at in Section 6.6.

6.3 Blastocyst grading

Blastocysts used for this study were obtained by the *in vitro* culture of spare IVF embryos. Spare embryos are handed over for research at the cleavage stage on day 2. At this time (around 40-46h post-insemination) they are usually at the 4-6 cell stage. Cell division continues *in vitro* with the embryos reaching the 8 cell stage (64-70h post-insemination) on day 3. The morula stage (12-16 cells) is normally reached on day 4 (88-94h post-insemination). By this time the embryo appears like a ball of cells although unlike mouse embryos, compaction with an indistinct cell outline is not always observed. Early cavitation occurs about 12-18 hours later gradually expanding to occupy a large volume. The blastocyst stage where the cavity is surrounded by a layer of trophectoderm and lined on one side by a distinct inner cell mass is normally reached on day 5 or 6 of culture. On further *in vitro* culture, the cavity expands (day 6 or 7) further, accompanied by the thinning of the zona pellucida. This results in an opening in the zona through which the blastocyst emerges or hatches (day 7 or 8). The cavity of the hatched blastocyst normally collapses within a day or two. Sometimes however, the blastocoel cavity may collapse even before hatching has taken place. The formation, expansion hatching and collapse of the blastocyst is represented in Plate C.

The pace of development in the field of *in vitro* embryo culture during the first half of this decade has been very rapid. This has resulted in techniques such as co-culture that allow more than 60% of embryos to

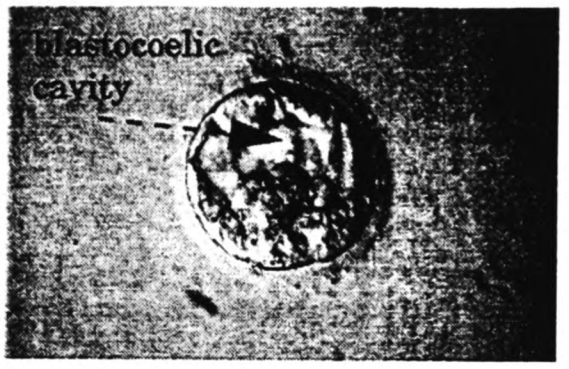
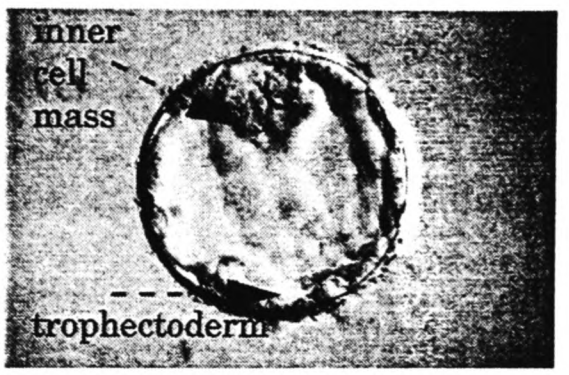
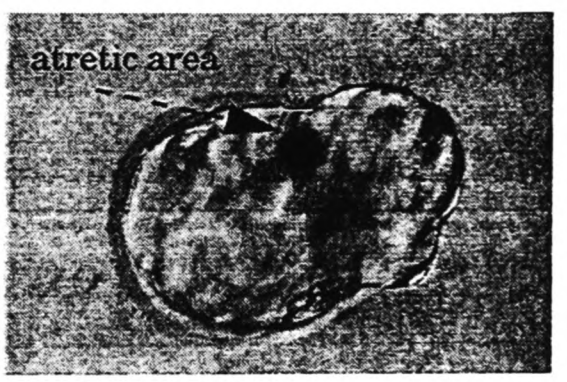
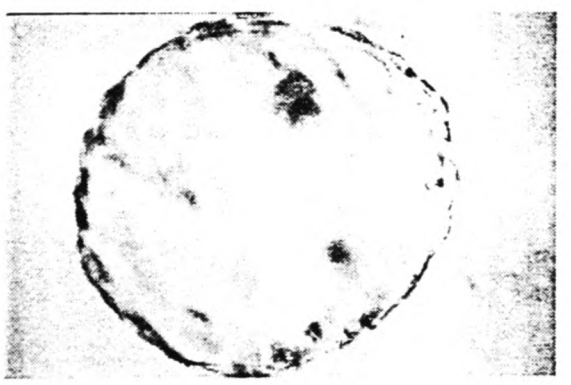
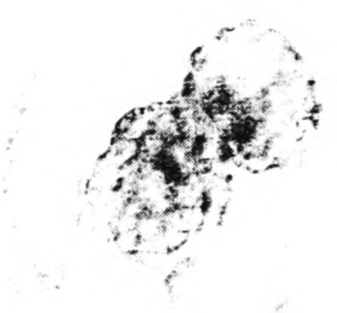
blastocyst stages		examples of features observed
early blastocyst		
expanding blastocyst		day of blastocyst formation cavity expanded day of expansion
hatching blastocyst		hatched day of hatching
hatched		collapse day of collapse atretic areas vacuoles
collapsing; collapse could occur before or after haching		

Plate C. Blastocysts observable at different stages on *in vitro* culture and examples of some features

develop *in vitro* to the blastocyst stage (Menezo *et al*, 1990; Bongso *et al*, 1994; Ciriminna *et al*, 1995) in contrast to earlier figures of between 15-30% (Lopata and Hay, 1989; Bolton *et al*, 1989; Dokras *et al*, 1991). The availability of an increasing number of human blastocysts has naturally led to the opening of many avenues of application. Firstly, there is a renewal of interest in human IVF transfers at the blastocyst stage to allow a synchronisation of the embryo's stage of development with that of the endometrium as happens *in vivo*. Secondly, blastocyst freezing with the reported advantages of higher survival rates when compared with freezing at earlier stages (Fehilly *et al*, 1985, Cohen *et al*, 1986, Freitas *et al*, 1994) is becoming increasingly popular with IVF Units. Thirdly, it allows for the possibility of trophectoderm biopsy at the blastocyst stage as a means of pre-implantation genetic diagnosis (Dokras *et al*, 1990). Finally, it promotes the development of blastocyst stage implantation models that could provide insight into implantation failure and other forms of infertility.

While the increase in the percentage of embryos proceeding to the blastocyst stage in culture with its potential applications is encouraging, safeguards are required to ensure that this does not take place at the cost of lowering blastocyst quality. This highlights the need to identify reliable markers to assess the quality of blastocysts developed under varying culture conditions. Such markers could also prove useful in the selection of 'good' quality blastocysts for transfer.

6.3.1 Blastocyst grading scheme of Dokras *et al* (1993)

Until recently, research into the assessment of human blastocyst quality has occupied the back-burner. Few markers of quality have been investigated. The biochemical marker, hCG, found to correlate with embryo quality is only detectable overall, day 8 onwards. (Dokras *et al*, 1991). While useful to assess blastocyst quality, it is not practically useful to allow selection of blastocysts for transfer, which usually takes place on day 5 or 6. Only one study reports on the development of a method which could be used to classify blastocysts of differing developmental potential at an earlier stage. The grading scheme developed at the Oxford IVF Unit classifies blastocysts into three grades based on their morphology (Dokras *et al*, 1993). The rules for grading have been extracted from experience of *in vitro* culture up to and beyond the blastocyst stage and are subjective. The principle guiding the grading is the same as that used in the methods of embryo grading described in Section 2.1.1. Blastocysts are classified into 3 grades based on certain pre-defined combinations of morphological features related to the blastocoel cavity, atretic areas and vacuoles. Plate C shows the blastocyst and the course of its development as well as outlines some morphological features that may be observed.

The classification of Blastocysts into 3 morphological grades uses some of the features mentioned in Plate C. The 3 grades of blastocysts are shown in Plate D. Briefly, Grade 1 (**BG1**) blastocysts are characterised by an initially eccentric blastocoel progressing to a well expanded cavity lined by



BG1



BG2

(initially vacuolated; subsequently appearing similar to BG1)



BG3

Plate D. Blastocyst grades in the grading scheme of Dokras *et al* (1993)

a distinct inner cell mass (ICM) region and trophectoderm layer. Grade 2 (**BG2**) blastocysts are similar in appearance to **BG1** but are preceded by a transitional phase showing single or multiple vacuoles with sharp borders at the morula stage. Grade 3 (**BG3**) are blastocysts with or without initial vacuolation that show several degenerative foci in the ICM on the day of formation and an incompletely expanded cavity which collapses within 24 hours.

In the absence of practically useful biochemical markers of blastocyst grading, this morphological grading scheme is an attractive prospect. It is non-invasive and easily performed. It was also supported by observations which showed that a higher incidence of hatching and growth was observed in Grade 1 and 2 blastocysts as compared to the poorer quality Grade 3 blastocysts.

6.3.2 Drawbacks of the Dokras grading scheme

Some of the drawbacks of the grading scheme are as follows:

1. **Missing grades:** Grades are allotted to blastocysts based on certain pre-decided combinations of morphological features. In the interest of clarity, Table 6.1 converts the grades described in Dokras *et al* (1993) to a tabular form. While not all the combinations in the table may be biologically possible, blastocysts showing some of the combinations not accounted for in the Dokras scheme do develop. This leaves such blastocysts unclassified (see Table 6.1) necessitating their capricious

allotment to one of the 3 grades. This leads to an inconsistent classification system.

Table 6.1 Blastocyst grading scheme of Dokras *et al* (1993) reduced to a tabular form

Vacuoles + / −	Atretic areas + / −	Expanded cavity + / −	collapse within 24h + / −	Blastocyst Grade
+	+	+	+	?
+	+	+	−	?
+	+	−	+	?
+	+	−	−	?
+	−	+	+	BG2
+	−	+	−	BG2
+	−	−	+	?
+	−	−	−	?
−	+	+	+	?
−	+	+	−	?
−	+	−	+	BG3
−	+	−	−	?
−	−	+	+	BG1
−	−	+	−	BG1
−	−	−	+	?
−	−	−	−	?

+ = yes
− = no

2. **Pre-defined grades:** In this grading system blastocysts are allotted to grades based on pre-defined combinations of features. There is no room for identifying any feature combinations that might actually show a stronger correlation with blastocyst quality.

3. **Prospective approach:** In the grading scheme of Dokras *et al* (1993) blastocyst grades are first allotted based on morphology. The validity and ability of the grades to distinguish blastocysts of differing quality is then assessed by measuring objective quality markers (e.g. hCG, nuclei

counts) for blastocysts belonging to differing grades. The correlation of levels of the markers with blastocysts of different morphological grades is investigated. If a correlation is identified, it provides support to the morphological grading scheme (Dokras *et al*, 1993). If however, no correlation is identified, there is no way of using the information provided by such a marker to improve the grading scheme so that it captures the properties of the marker (e.g. pregnancy specific beta-1 glycoprotein - SP1, immunosuppressive activity in Appendix II). Prospective validation thus may not be very beneficial. In contrast to this approach, a retrospective approach to grading (described below) would be more profitable.

6.3.3 Alternate grading scheme using decision trees

A method of grading blastocysts using decision trees, which overcomes some limitations of the scheme of Dokras *et al* (1993) is described here. In contrast to the prospective approach above, this is a retrospective approach in which the combinations of morphological characteristics of the blastocysts found to correlate with known markers of blastocyst quality (e.g., hCG or other secreted factors) are grouped together to define the grades. The grading is thus based on the objective proven quality of the blastocyst. This approach is depicted in Figure 6.1 where it is contrasted against the prospective approach.

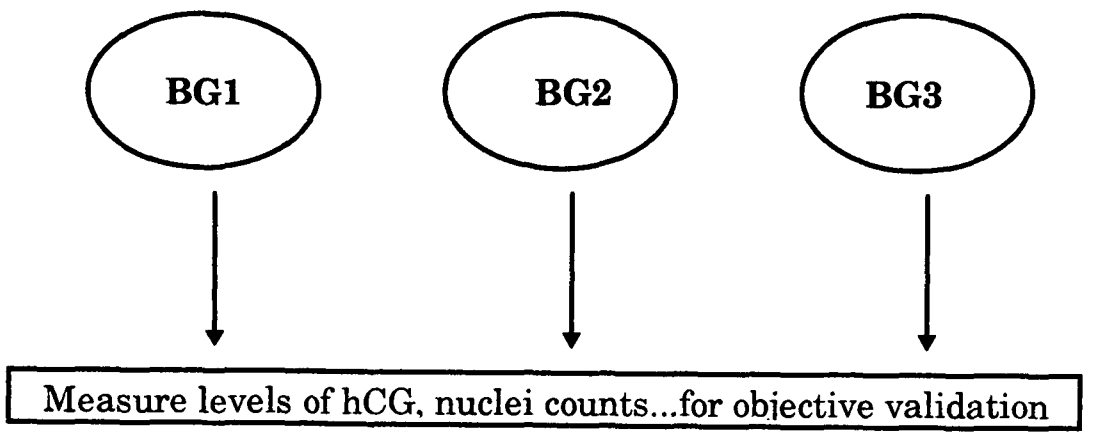


Figure 6.1 Prospective and retrospective approaches to grading

A simple application of the decision tree technique to grade blastocysts into one of two grades, 'good' and 'bad' is outlined here. 'Good' is taken to indicate blastocysts secreting high hCG levels (>45 IU/ml explained in Section 6.4.1) and 'bad' to indicate blastocysts secreting low hCG levels (<45IU/ml). The morphological features associated with blastocysts belonging to the two groups are then identified using decision tree analysis. At this point, it is important to understand the implication of the classification 'good' as defined here. It implies that if accurate rules of

morphological features characterising 'good' blastocysts can be obtained, then we can expect blastocysts satisfying these rules to show high hCG secretion. Similarly blastocysts satisfying rules characterising 'bad' blastocysts can be expected to secrete low levels of detectable hCG.

6.4 Materials and Methods

Figure 6.2 gives a broad outline of the steps followed in this study. Each step is elaborated further in relevant sections below.

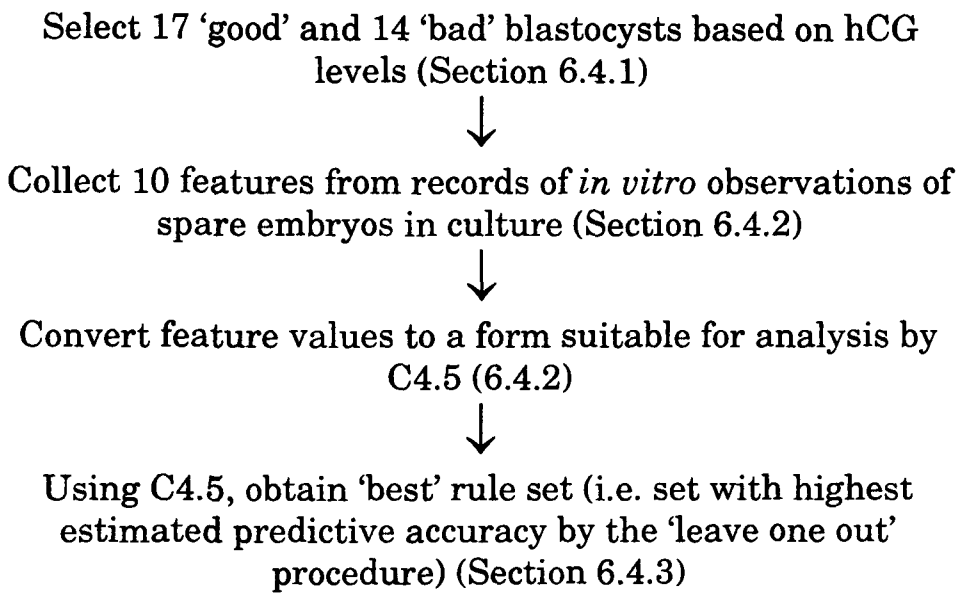


Figure 6.2 Outline of methods

6.4.1 Blastocyst selection.

The 31 spare embryos from a previous study (Dokras *et al*, 1993) that were also considered in the pilot study in Chapter 4 were included in the analysis here (Section 4.3.2). These were embryos that had progressed to the blastocyst stage and had been cultured *in vitro* until they were 14 days. They were morphologically assessed every 24 hours before being transferred to fresh medium (for details, see Sections 3.3.2 and 3.3.3). The

daily levels of hCG secreted by these embryos, in the surrounding medium, measured from day 3[1] to day 14 were available from the previous study.

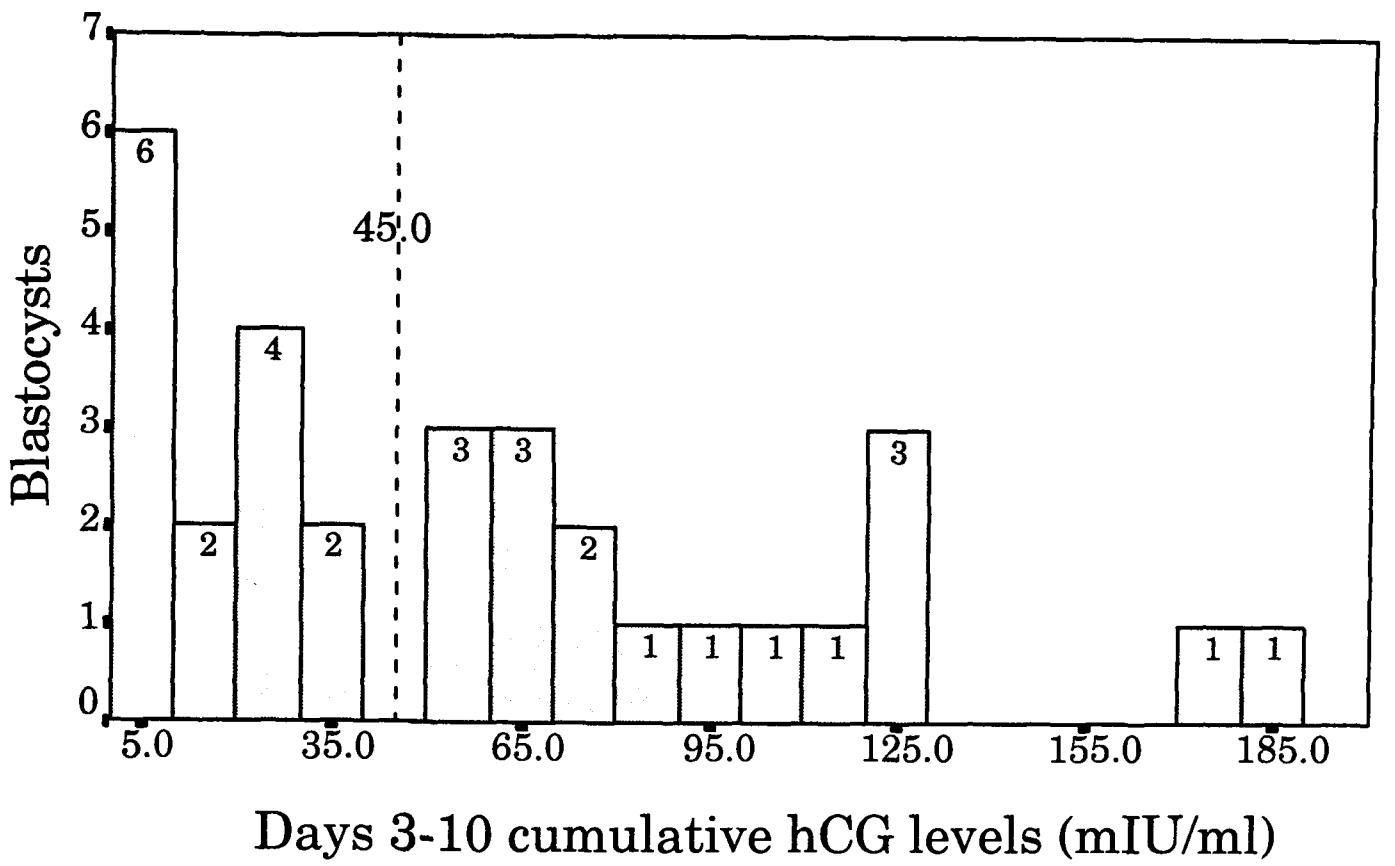
hCG secretion, accepted widely as a marker of embryo quality (Confino *et al*, 1986 and Dokras *et al*, 1993), was chosen as the criterion to classify blastocysts as 'good' or 'bad' (Section 4.3.2). Day 10 was taken as the cut-off and the cumulative hCG secretion from day 3-day 10 was considered to be a reflection of the quality of the blastocyst. The reasons for this have been described in detail in Chapter 4 (4.3.2). Figure 6.2 below reproduces the histogram of cumulative levels of hCG secretion showing the demarcation (45mIU/ml) between the secretion of high and low levels.

The aim of the analysis here was to identify the morphological features of the blastocysts associated with 'high' and 'low' levels of hCG secretion. For the purpose of this study therefore, the 17 high (>45 IU/ml) hCG secretors were classified as 'good' and 14 low (<45IU/ml) hCG secretors as 'bad' blastocysts.

---

[1] Medium collected from embryos when allotted for research on day 2 contains factors secreted by the embryo over the previous 24 hours since the detection of fertilisation and is therefore referred to as day 1, when referring to embryo-derived products. Similarly medium collected on day 3 of culture is designated day 2 and so on. This holds good for the sense in which 'day' is used when referring to hCG secretion in this chapter as well as IS activity and SP1 levels in Appendix B

Figure 6.3 Histogram showing the distribution of blastocysts according to cumulative (days 3-10) hCG values



6.4.2 Feature selection, collection and conversion

Spare embryos were morphologically assessed every 24 hours up to the blastocyst stage using a low power stereo dissecting microscope. Blastocysts were then assessed 24 hourly up to day 14 using a high power Wilovert inverted phase contrast microscope. A record of the morphological features observed was maintained as part of a previous study carried out by A. Dokras. For the analysis here, these features were extracted from the records and then converted to a form acceptable for analysis by C4.5. Table 6.2 shows the morphological features included in this study together with nominal or ordinal level of measurement (Section 4.3.3) and numerical value or code used here.

Table 6.2 10 features selected for analysis

No	Feature	Range of values	Level	Numerical value or Code[2]
1	day of blastocyst formation	5-7	Ratio	same as original value[3]
2	expanded cavity	yes no	Nominal	same as original value
3	day of expansion	6.5-8	Ratio	same as original value[4] and 0 if the blastocyst did not expand
4	hatched	yes no	Nominal	same as original value
5	day of hatching	7-9	Ratio	same as original value and 0 if the blastocyst did not hatch
6	collapse	yes no	Nominal	same as original value
7	day of collapse	7-9	Ratio	same as original value and 0 if the cavity did not collapse[5]
8	atretic area[6]	yes no	Nominal	same as original value
9	vacuoles	yes no	Nominal	same as original value

[2] If an observation for any of the morphological features of any of the blastocysts was not recorded, following Quinlan (1993), the value was denoted by '?'.

[3] In cases recorded initially as early blastocyst and then progressing to the blastocyst stage, the day of blastocyst formation was taken to be that between the 2 events (for e.g. if the records showed day of early blastocyst formation as day 5 and blastocyst as day 6, then the day of blastocyst formation was taken to be 5.5)

[4] If a blastocyst was recorded as having formed on a particular day and then recorded as hatched, on a later date, the day of expansion was taken to be that between these 2 readings

[5] Most blastocysts collapsed either before or after hatching. The day of collapse was recorded as 0 for the few that still had an intact cavity on day 14

[6] It was not clear in the records, whether the observations referred to atretic areas restricted to the inner cell mass or included atretic areas seen distributed in the surrounding trophectoderm

10	first day of visualisation of vacuoles	4-6	Ratio	same as original value and 0 if no vacuoles were visualised

Table 6.3 shows the first few items of data in the final form ready to be used as the input for C4.5.

Table 6.3 Examples of data in the final form ready for analysis

	Value of feature for each blastocyst				
Blastocyst	Day of blastocyst formation	Expanded cavity	Day of expansion	...	Class
Blastocyst 1	7	No	0	...	bad
Blastocyst 2	6	No	0	...	bad
Blastocyst 3	5	Yes	6	...	good
...	...	...	...	...	
Blastocyst 'n'	6	Yes	7	...	good

6.4.3 Obtaining rules

The data for the 31 blastocysts was used by C4.5 to construct rules expressing the pattern of morphological features characterising blastocysts secreting high or low levels of hCG. C4.5 was used to construct the 'best' rule set, i.e. the rule set with the highest estimated predicative accuracy (see Section 3.5.4). The cross validation procedure of 'leave one out' was followed to obtain the estimate of predictive accuracy (see Section 3.5.4).

6.5 Results

First the rules obtained are shown. Then the 2×2 table representing the performance from 'leave one out' cross validation, its χ^2 value and estimate of predictive accuracy are given

Rules

Rule 1

If
expanded cavity = no and
day of collapse ≤ 8
Then
class = **bad**

Rule 2

If
atretic areas = yes and
day of collapse ≤ 8
Then
Class = **bad**

Rule 3

If
day of collapse > 8
Then
Class = **good**

Rule 4

If
atretic areas = no and
expanded cavity = yes
Then
Class = **good**

Rule 5 (default rule)

If
none of the above
Then
Class = **bad**

'leave one out' cross validation 2×2 table

		Actual class		
		good	bad	
Predicted class	good	14	6	20
	bad	3	8	11
		17	14	31

χ^2 value = 3.6

Estimate of predictive accuracy of the rule set = 71% (± standard deviation of 8.2%)

6.6 Discussion and Conclusions

Data related to 31 blastocysts was analysed to develop rules that expressed the morphological features characterising 'good' and 'bad' blastocysts, i.e. blastocysts secreting high or low levels of hCG. The χ^2 value of 3.6 indicates that there is only a 5.8% probability that the pattern identified by the rule set is one that does not really exist, i.e. that it is a chance effect. Estimated predictive accuracy is defined as previously in this dissertation.

The rules obtained here could be used prospectively to classify blastocysts into one of 2 classes 'good' or 'bad'. Rule 1 classifying blastocysts in which the cavity failed to expand and collapsed early, as 'bad', indicates an impaired functioning of the trophectoderm cells. This is because water transport across the trophectoderm occurs passively as a result of the

active transport of ions by the trophectoderm cells (Biggers *et al*, 1988 in rabbit; Watson, 1992 in mouse; Brison and Leese, 1993 in rat blastocysts). Such impaired functioning is of course reflected in the relatively low (<45IU/ml) levels of hCG that can be expected to be secreted by blastocysts observed to have the above characteristics. Similarly in Rule 2 the presence of atretic areas associated with 'bad' blastocysts, is reflected by the low levels of secretion of hCG by such 'bad' blastocysts.

The presence of a well expanded cavity and the persistence of the cavity until after the 8th day of culture in the 'good' blastocysts as identified in the rule set (Rules 3 and 4), indicates the well-being of the trophectoderm cells that were able to successfully maintain the transepithelial flux that controls the transfer of fluid into the blastocoelic cavity. This state of trophectoderm cells is reflected in the ability of the blastocysts to secrete high (>45IU/ml) levels of hCG, as is the absence of atretic cells.

The rules do not identify hatching as important in differentiating blastocysts with differing capacities of hCG secretion. Although this appears unusual at first, it is not surprising given the findings of Dokras *et al* (1991) who reported on the absence of an overall difference in the hCG secretion by hatched blastocysts and those that remained intact within the zona. Also the presence or absence of vacuoles bears no relationship to the expected level of hCG secretion. Again Dokras *et al* (1993) showed that the Grade 2 blastocysts (characterised by vacuoles)

showed no significant difference from the Grade 1 (no vacuoles) in their ability to expand, hatch or secrete hCG.

Based on the morphological characteristics associated with a blastocyst, it is possible to use these rules to predict if it would secrete high or low levels of hCG. As compared to the morphological grading scheme of Dokras *et al*, the rules obtained here are simple and well defined. If a sufficiently large number of blastocysts were available, this method could even be used to classify blastocysts into more than 2 classes. The morphological features commonly characterising blastocysts secreting not just high or low but varying levels of hCG along a scale could be investigated. The analysis could also be extended further such that the morphological features characterising blastocysts capable of secreting particular levels of more than one substance could be investigated, e.g. capable of secreting high levels of hCG as well as high levels of SP1 or low levels of hCG as well as SP1. The resulting rules would then have captured the combination of features characterising blastocysts capable of secreting high hCG as well as high SP1 or low hCG as well as low SP1.

The rules could also be used to assess and select 'good' blastocysts for transfer or select for cryopreservation the blastocysts most likely to survive freezing. For this however, features recorded at an earlier stage of blastocyst development (before day 6 or 7) would have to be included in the analysis.

Although the study reported here has the drawback of a very small sample of 31 blastocysts and an arbitrary cut-off between low and high hCG values, it serves to illustrate the applicability of the ideas developed in this dissertation to other areas of IVF.

6.7 Summary

This chapter has explored the use of decision tree techniques for the grading of *in vitro* cultured blastocysts of differing quality. Simple rules expressing morphological features characterising blastocysts capable of secreting high or low levels of hCG have been obtained. The main aim here was to illustrate *via* an example the potential for the application of ideas developed in the dissertation to other areas of IVF.

Chapter Seven. General Discussion And Conclusion

7.1 SUMMARY 163

7.1.1 Main Aspects 163

7.1.2 Contributions Of The Dissertation 165

7.2 FUTURE WORK: EXTENSIONS TO THE MAIN IDEA 169

7.3 FUTURE WORK: OTHER APPLICATIONS 171

7.4 END NOTE 172

7. GENERAL DISCUSSION AND CONCLUSION

This dissertation has been concerned with investigating the relationships of embryo-related features to the developmental potential of human IVF embryos. The 'raw materials' for this investigation have been data records extracted from IVF patient files, and an analysis technique called 'decision trees'. This chapter summarises the nature, course and results of the ensuing examination. It also proposes extensions to broaden the scope of the inquiry to include maternal and sperm factors and suggests ways by which methods adopted here could be used in other areas of IVF.

7.1 Summary

7.1.1 Main aspects

Despite extensive advances in IVF, allowing the observation and recording of data related to various stages of the IVF process, the studies attempting to extract useful information from such data have suffered from many limitations. The aim of the study here was to provide a clearer understanding of the relationship between a large number of embryo, oocyte and associated follicular features recorded in IVF patient files and the development potential of human IVF embryos on transfer. The technique called 'decision tree' analysis, developed in Machine Learning in Computer Science was used as a tool towards this end. Their application to IVF data was examined over the course of three experimental studies.

The feasibility of using the decision tree technique to analyse IVF data, was first explored in a pilot study. The data comprised of *in vivo* transferred as well as *in vitro* cultured embryos. The relationship between 15 embryo-related features and the developmental potential was explored and found to be expressible as simple, reasonably accurate rules.

Given the promising nature of the results of the pilot study, a larger study, which forms the core of the dissertation, was conducted. Data included batches of three embryos transferred to patients and resulting in either a TH outcome or a negative pregnancy test (NTH). The unit of analysis here was a batch of three embryos rather than individual embryos. The relationship between 53 embryo-related features and the 'TH' or 'NTH' developmental potential was identified. There do not appear to be any other reports in the literature of an investigation involving such a large number of features extending over various stages of the IVF process (follicle up to cleaving embryo).

The pilot and batch studies were followed by a relatively modest study exploring the use of similar techniques to grade blastocysts. The rules identifying the morphological features characterising blastocysts of differing quality were simple and a biological explanation for these associations could be offered. The main role of this last experimental study was to indicate that the methods developed over the course of the pilot and batch studies were not necessarily restricted to the analysis of

embryos at an early stage. In fact, they could be used to clarify many other relationships in other areas of IVF.

7.1.2 Contributions of the dissertation

This dissertation has provided a meeting ground for state-of-the-art Machine Learning Analysis techniques with invaluable IVF data records. Out of this convergence, have emerged insights into some of the factors influencing the developmental process of the human embryo as well as observations of relevance to the embryo selection procedure employed by the IVF Unit in Oxford.

Main findings of relevance to the relationship between the features investigated and the developmental potential are as follows:

1. Of the 53 features investigated, only 4 namely, embryo grade, cell number, follicle size and follicular fluid volume were identified as having an important relationship with the embryos' developmental potential.

2. The identification of follicle size and follicular fluid volume as important factors confirms that the developmental potential of the embryo is determined as early as the stage of the unfertilised oocyte. The values for these features point to a 'oocyte maturity window'.

3. The results demonstrate that the developmental potential is not directed by a single factor, but is influenced by a number of factors acting in combination. For example, embryos originating from oocytes

of similar maturity could vary in their developmental potential depending on the influence of other concomitant factors. Single features should not therefore be used to pass judgement on the developmental potential of the embryo.

4. By virtue of exclusion, the study identifies a large number features that do not appear to have an important association with the developmental potential. These include features related to the appearance of the oocyte, oocyte grade, visibility of nuclei within the blastomeres, the cumulus dish attachment, the intercellular regularity and fragmentation in the morning on the day of transfer.

5. Although the embryo, oocyte and associated follicular features of a batch may suggest a good developmental potential, the embryos may still fail to thrive. The influence of other factors, possibly maternal, on transfer could be responsible for the failure of pregnancy.

The results reported here are within the specific context of one IVF laboratory. It is hoped that the method of analysis will be replicated and evaluated in other laboratories.

Comments worthy of note in relation to the embryo selection procedure employed in the Oxford IVF Unit are:

1. The identification of the embryologists grading system and follicle size amongst the important features provides, for the first time, an unbiased verification of assessment methods used here. Furthermore, this study is capable of contributing to the assessment procedure as it makes

quantitatively precise the values of embryo grade and follicle size that indicate the 'take home baby' or 'no take home baby' development potential of a batch of embryos.

2. The detailed investigation of the embryo grade (given after the assessment of 5 component features) indicates that the essence of the grade 1 embryos is captured by the rate of cleavage, as reflected by the cell number of the embryo on the morning of the day of transfer. The study therefore suggests this feature as a viable alternative to embryo grade, in essence dropping fragmentation and irregularity as unimportant to the outcome at this grade = 1 level.

3. Two additional features namely embryo cell number on the morning of the day of transfer (mentioned in 2 above) and ff volume have been identified as important and therefore worthy of further investigation.

4. The results point to the relative unimportance of certain other features, such as the darkness of the corona and visibility of nuclei in the blastomeres, currently used in the selection process but not identified in the present study as bearing any correlation with the developmental potential. Similarly, the results confirm the unimportance of some other features that are not directly used in the selection process. These include the amount and description of the oocyte cumulus, maturity of the corona, oocyte grade, cumulus dish attachment, the regularity, fragmentation and visibility of nuclei within blastomeres in the morning on the day of transfer.

The improved understanding of relationships could be useful in assisting embryologists in an informed selection of the 'best' embryos for transfer. With the current clinical trend to reduce the number of embryos transferred to two (Vauthier-Brouzes *et al*, 1994; Staessen *et al*, 1993), the right choice of embryos has become even more imperative than before. Information about the likely developmental outcome of the embryos transferred and thus a realistic estimate of the chance of success could additionally serve as an aid to counselling IVF patients.

Apart from the findings in the two areas mentioned above, a contribution of this dissertation is the introduction of the use of the decision tree technique, the inherent properties of which make it a valuable tool for the analysis of IVF data. With the proliferation of new assays for measuring embryo-derived products and maternal hormone levels, improved ultrasound methods as well as the increasing use of microscopes with a very high resolution capacity, the availability of new features is very likely to increase over the next decade. Given the inability of single morphological features or single substances measured in ECF to distinguish embryos of differing quality, decision tree techniques which are flexible and capable of accommodating a large number of features can be used to combine the information from a number of features towards identifying the good quality embryos.

The possible use of these techniques in exploring relationships at a later stage of embryo development has been indicated in the dissertation

through the development of a preliminary blastocyst grading scheme. The decision tree analysis technique is thus quite versatile with a potential for application to other areas in IVF.

7.2 Future work: extensions to the main idea

This dissertation mainly investigates the relationships between embryo-related features and development. Embryo development however, is also influenced by the quality of the sperm of origin as well as the maternal factors. An investigation into the relationship between these features and developmental potential is a natural extension to the main idea developed here.

Although the features associated with an embryo may be suggestive of a very good developmental potential, a hostile maternal environment could prevent it from progressing to a pregnancy. For example an embryo identified as having a good developmental potential may result in a successful pregnancy in a younger woman but could fail to implant when transferred to a woman who is older (Feldberg *et al*, 1990). The age of the mother, the cause of infertility, uterine structure, the drugs used, levels of reproductive hormones all affect the receptivity of the uterus to the implanting embryo and thus influence the developmental potential. It is therefore worthwhile investigating the extent to which each of these features influences development. Such information could then be used for example to design drug protocols so as to create the most favourable conditions for embryo development for individual patients. Another

application could be in suggesting the postponement of the embryo transfer to another cycle in order to increase the likelihood of success. Such a situation could arise if, in a given cycle, embryo features were identified by the rules as indicating a 'good' developmental potential, but maternal features were not. The embryo could then be frozen until a cycle with favourable features was achieved. Caution must be exercised to ensure that the conditions under which rules are applied are broadly similar to those under which they were obtained e.g. if using rules to guide selection of cryopreserved embryos, the effect of cryopreservation needs to have been accounted for by training on similar data.

Similarly sperm parameters such as motility, morphology, volume, and concentration appear to influence the developmental potential of the embryos (Kruger *et al*, 1986 and 1987; Oehninger *et al*, 1988; Enginsu *et al*, 1992; De Geyter *et al*, 1992). The view that once fertilisation and cleavage stage embryos have been achieved, no matter what the quality of sperm used, the chances of pregnancy are similar (Van Uem *et al*, 1985; Tournaye *et al*, 1992 and Acosta *et al*, 1986), is no longer tenable. The influence of different sperm parameters in relation to the maternal and embryo features thus requires investigation.

Having investigated in different studies the relationship between embryo-related, maternal and sperm features and the developmental potential, the next step would be to analyse all these features together. The aim would be to study the interactions between maternal, sperm and embryo-

related features and obtain rules that indicate the impact of differing combinations of these features on the ability of an embryo to establish a successful pregnancy. Taken together, the studies could form the basis of an embryo selection procedure that objectively accounts for all three sets of features.

Before such rules are used to assist embryo selection, the size of the data sample used should be increased from that used here to give the rules greater statistical validity. The issue of 'costs' also needs to be taken into account when obtaining the rules.

7.3 Future work: other applications

The ideas developed in this dissertation are not necessarily restricted to the analysis of embryos at an early stage. In fact, they could be used to clarify many other relationships in other areas of IVF. To consider but a few, the technique can be used to investigate the relationship between various oocyte and embryo features and the ability of an embryo to survive cryopreservation. Such information can be used to direct the choice of embryos to be cryopreserved. Similarly, the technique can be used to explore the relationship between the components of the cryopreservant or culture media and embryo survival. This could be utilised towards improving embryo culture and cryopreservation conditions. The blastocyst grading, briefly dealt with in Chapter 6 and the areas mentioned here indicate the wide scope of this method and the open-ended nature of the research.

7.4 End note

This work is viewed primarily as a scientific investigation into the relationships between embryo-related features and development. The investigation needs to be extended along the lines discussed earlier and its limitations addressed before the findings are ready for clinical uptake. Moreover, translating the results obtained into clinical practice can only be done in consultation with expert clinicians. The information obtained should be viewed as something to be used by the experts, rather than instead of them.

Appendix A.

APPENDIX A

53 features included in the analysis in Chapter 5 and their values as recorded in IVF treatment files together with the level of measurement and numerical value or code.

N	Feature	Range of values	Level	Number or code
1	follicle size at oocyte recovery (day 0)	10 - 38 mm	ratio	same as original value
2	ff volume at oocyte recovery (day 0)	0.5 - 9.0 ml	ratio	same
3	oocyte cumulus amount (day 0)	very little amount medium lots	ordinal	1 2 3
4	oocyte cumulus description (day 0)	dark (immature)[1] slightly dark dark (just post mature) very slightly post mature slightly post mature post mature	ordinal	1 1 1 2 3 4
5	oocyte corona maturity (day 0)	very immature immature mature	ordinal	1 2 3
6	oocyte corona description (day 0)	okay slightly dark dark borderline very dark	ordinal	1 2 3 4 5
7	darkness of the ooplasm (day 0)	none slightly dark dark very dark	ordinal	0 1 2 3

[1]The IVF records use the same keyword "dark" in two different contexts. It is impossible to distinguish between darkness related to immaturity, and that related to postmaturity. To simplify this, all dark values are given the same numerical code.

8	granularity of the ooplasm (day 0)	none slightly granular granular very granular	ordinal	0 1 2 3
9	elongation of the ooplasm (day 0)	none slightly elongated elongated very elongated	ordinal	0 1 2 3
10	shrunken appearance of the ooplasm (day 0)	none slightly shrunken shrunken very shrunken	ordinal	0 1 2 3
11	vesicles in the ooplasm (day 0)	numerical	ratio	same
12	refractile bodies in the ooplasm (day 0)	numerical	ratio	same
13	misshapen ooplasm (day 0)	none slightly misshapen misshapen very misshapen	ordinal	0 1 2 3
14	thickness of the zona pellucida (day 0)	none thick very thick	ordinal	0 1 2
15	elongation of the zona pellucida (day 0)	none slightly elongated elongated very elongated	ordinal	0 1 2 3
16	egg grade	post mature mature borderline immature immature	ordinal	1 2 3 4
17	darkness of the cytoplasm (fertilisation check, day 1)	none slightly dark dark very dark	ordinal	0 1 2 3
18	granularity of the cytoplasm (fertilisation check, day 1)	none slightly granular granular very granular	ordinal	0 1 2 3
19	elongation of the cytoplasm (fertilisation check, day 1)	none slightly elongated elongated very elongated	ordinal	0 1 2 3

20	shrunken appearance of the cytoplasm (fertilisation check, day 1)	none slightly shrunken shrunken very shrunken	ordinal	0 1 2 3
21	vesicles in the cytoplasm (fertilisation check, day 1)	numerical	ratio	same
22	refractile bodies in the cytoplasm (fertilisation check, day 1)	numerical	ratio	same
23	misshapen cytoplasm (fertilisation check, day 1)	none slightly misshapen misshapen very misshapen	ordinal	0 1 2 3
24	thickness of the zona pellucida (fertilisation check, day 1)	none thick very thick	ordinal	0 1 2
25	elongation of the zona pellucida (fertilisation check, day 1)	none slightly elongated elongated very elongated	ordinal	0 1 2 3
26	cumulus dish attachment	massive lots moderate little none	ordinal	1 2 3 4 5
27	number of cells (check 1, day 2)	1 - 10	ratio	same
28	regularity of cells (check 1, day 2)	not applicable[2] regular slightly irregular irregular very irregular	ordinal	? 1 2 3 4

[2]Regularity across cells and visibility of nuclei are not applicable at the pronuclear and syngamy stage. Following Quinlan (1993), values that are not applicable are denoted by '?'

29	fragmentation (check 1, day 2)	no fragments	ordinal	1
		few tiny fragments		2
		few small fragments		3
		few fragments		4
		fragments		5
		lots of fragments		6
30	visibility of nuclei in the cells (check 1, day 2)	not applicable[3]	nominal	?
		no		no
		yes		yes
31	darkness of the cytoplasm (check 1, day 2)	none	ordinal	0
		slightly dark		1
		dark		2
		very dark		3
32	granularity of the cytoplasm (check 1, day 2)	none	ordinal	0
		slightly granular		1
		granular		2
		very granular		3
33	elongation of the cytoplasm (check 1, day 2)	none	ordinal	0
		slightly elongated		1
		elongated		2
		very elongated		3
34	shrunken appearance of the cytoplasm (check 1, day 2)	none	ordinal	0
		slightly shrunken		1
		shrunken		2
		very shrunken		3
35	vesicles in the cytoplasm (check 1, day 2)	numerical	ratio	same
36	refractile bodies in the cytoplasm (check 1, day 2)	numerical	ratio	same
37	misshapen cytoplasm (check 1, day 2)	none	ordinal	0
		slightly misshapen		1
		misshapen		2
		very misshapen		3
38	thickness of the zona pellucida (check 1, day 2)	none	ordinal	0
		thick		1
		very thick		2
39	elongation of the zona pellucida (check 1, day 2)	none	ordinal	0
		slightly elongated		1
		elongated		2
		very elongated		3

[3] See explanation in footnote 2

40	number of cells (check 2, day 2)	1 - 10	ratio	same
41	regularity of cells (check 2, day 2)	not applicable	ordinal	?
		regular		1
		slightly irregular		2
		irregular		3
		very irregular		4
42	fragmentation (check 2, day 2)	no fragments	ordinal	1
		few tiny fragments		2
		few small fragments		3
		few fragments		4
		fragments		5
		lots of fragments		6
43	visibility of nuclei in the cells (check 2, day 2)	not applicable	nominal	?
		no		no
		yes		yes
44	darkness of the cytoplasm (check 2, day 2)	none	ordinal	0
		slightly dark		1
		dark		2
		very dark		3
45	granularity of the cytoplasm (check 2, day 2)	none	ordinal	0
		slightly granular		1
		granular		2
		very granular		3
46	elongation of the cytoplasm (check 2, day 2)	none	ordinal	0
		slightly elongated		1
		elongated		2
		very elongated		3
47	shrunken appearance of the cytoplasm (check 2, day 2)	none	ordinal	0
		slightly shrunken		1
		shrunken		2
		very shrunken		3
48	vesicles in the cytoplasm (check 2, day 2)	numerical	ratio	same
49	refractile bodies in the cytoplasm (check 2, day 2)	numerical	ratio	same
50	misshapen cytoplasm (check 2, day 2)	none	ordinal	0
		slightly misshapen		1
		misshapen		2
		very misshapen		3
51	thickness of the zona pellucida (check 2, day 2)	none	ordinal	0
		thick		1
		very thick		2

52	elongation of the zona pellucida (check 2, day 2)	none	ordinal	0
		slightly elongated		1
		elongated		2
		very elongated		3
53	embryo grade (subjective scoring of embryologist)	grade 1	ordinal	1
		grade 2		2
		grade 3		3
		grade 4		4
		grade 5		5

Appendix B. Assessment of IS Activity and SP1 Levels in Embryo Culture Fluid

B.1 INTRODUCTION 181

B.2 ORGANISATION OF THIS APPENDIX 181

B.3 IS ACTIVITY AND SP1 SECRETION: BACKGROUND 182

B.4 IS ACTIVITY: MATERIALS AND METHODS 185

B.4.1 Culture Medium Samples 185

B.4.2 Principle Of Method Used To Assess Is Activity: 186

B.4.3 Is Assay: Con A Slr 186

B.4.4 Calculation Of Is Activity 188

B.4.5 Is Activity In Fresh Versus Frozen Ecf 188

B.5 SP1: MATERIALS AND METHODS 189

B.5.1 Culture Medium Samples 189

B.5.2 Sp1 Assay Principle 189

B.5.3 Elisa 190

B.6 NOTE ON STATISTICAL CALCULATIONS AND TESTS USED 191

B.7 IS ACTIVITY: RESULTS AND DISCUSSION 192

B.7.1 Results 192

B.7.2 Discussion 197

B.8 SP1 SECRETION: RESULTS AND DISCUSSION 198

B.8.1 Results 198

B.8.2 Discussion 203

B.9 CONCLUSIONS 206

B.10 SUMMARY 206

Appendix B Assessment of IS activity and SP1 levels in Embryo culture fluid

B.1 Introduction

Chapter 6 has discussed the need for a reliable blastocyst grading scheme as well as for objective markers of blastocyst quality. It outlines the morphological grading of Dokras *et al* (1993) that was supported by measurements of hCG secretion and counts of nuclei within the blastocysts. The study outlined in this Appendix investigates the association between the Dokras grading scheme and other objective markers, namely, immunosuppressive (IS) activity and levels of pregnancy-specific beta-1 glycoprotein (SP1) in the ECF.

B.2 Organisation of this Appendix

Relevant studies investigating the relationship between IS activity (or SP1) and embryo or blastocyst quality are reviewed briefly in Section B.3 below. Section B.4 outlines the Materials and Methods used in the measurement of IS activity. This includes details related to the culture medium samples, the assay used to measure IS activity and the comparison of such activity in fresh medium with frozen medium. Section B.5 presents the Materials and Methods used to measure levels of SP1, including the collection of culture medium samples, principle and details of the ELISA utilised. Materials and Methods sections are followed by an

explanation of the statistical tests used to investigate the association between the levels of IS activity (and SP1) in ECF and the morphological grade of blastocysts (Section B.6). Results are presented and discussed in Section B.7 for IS activity and Section B.8 for SP1. Conclusions that can be drawn from the Results are in Section B.9.

B.3 IS activity and SP1 secretion: Background

IS activity and SP1 were chosen here for investigation as previous research, outlined below, suggests the possibility of a correlation of their levels in ECF and maternal serum, with embryo quality and subsequent foetal well -being.

IS activity

Research into embryo derived IS activity as a possible means of escaping the maternal immunological response to the semi-allogenic embryo, led a number of groups to investigate the ability of ECF to suppress lymphocyte proliferation in a mitogen induced lymphocyte proliferation reaction. This was considered as indirect evidence for the secretion of IS factor(s) by the embryo. Daya *et al* (1986) and Clark *et al* (1989) showed that 16% and 36% of patients respectively, who received at least one embryo associated with a suppressive activity, became pregnant. Similarly, Sheth *et al* (1991) claimed that the ECF from embryos that did and did not result in pregnancy showed significant differences in levels of both interleukin-1-alpha and IS activity. A correlation between IS and pregnancy was also

reported by Bose and Mahadevan (1990), Bose (1991) and Jones *et al* (1992).

IS factor(s) have been reported in the culture medium from rabbit, murine, bovine, ovine and porcine blastocysts (Dey *et al*, 1981; Van Vasselaer P *et al*, 1984; Mayumi *et al*, 1985, 1982; Fisher *et al,* 1985; Murray *et al*, 1987). There has been only one report of such a study performed on human blastocysts by Segars *et al* (1989) who were unable to detect any IS activity in the ECF of embryos that progressed to the morula stage and subsequently degenerated, or formed abnormal blastocysts. However, when the conditioned medium from two embryos that progressed to the blastocyst stage was sequentially analysed, a progressive increase in IS activity was detected. If IS activity does indeed have a role to play in the suppression of embryo rejection, this finding would be expected. *In vivo*, the embryo hatches from the surrounding zona pellucida around day 6 at the blastocyst stage and begins invading the uterine endometrium - any need for suppression therefore being higher at this point compared with earlier stages.

Based on the above findings, and the correlation between IS activity in ECF and pregnancy on transfer, it would be expected that good quality blastocysts would show a progressive rise in IS activity with increasing duration of culture, with high levels around the time of implantation. The poorer quality embryos and blastocysts with a lower potential for implantation, on the other hand, would not be expected to demonstrate

such a pattern. The IS activity in the 3 morphological grades of blastocysts (grades described in 6.2.1) over 12 days of culture was therefore assessed in order to investigate the association (if any) between morphological grading and pattern of IS activity.

SP1

Pregnancy-specific beta-1 glycoprotein (SP1), like hCG, is secreted by the trophoblast, and is detectable in the serum of pregnant women as early as day 7 after the luteinising hormone peak (Grudzinskas *et al,* 1977). Its levels are predictive of pregnancy outcome (Ahmed and Klopper, 1984). Later in pregnancy, serum SP1 levels can distinguish between normal singleton pregnancies and abnormal pregnancies such as recurrent spontaneous abortions, intra-uterine growth retardation, ectopic pregnancies, low birth weight and pre-term labour (Sinosich *et al,* 1985; Raman and Pawashe, 1986; Johnson *et al,* 1993a; Zhang *et al,* 1993; Johnson *et al,* 1993b). The exact role of SP1 remains unknown, but an immunosuppressive role in protecting the fetus from rejection by the maternal immune system has been postulated (Bischof, 1984).

As with IS activity, given the biological significance of SP1 suggested by its correlation with foeto-placental well-being, this study also sought to measure levels of SP1 in the embryo culture fluid (ECF) and investigate the association (if any) between levels of SP1 and the three morphological grades of blastocysts.

Vacuolated morulae (morulae that showed the presence of vacuoles which could occasionally be mistaken for early cavitation) were also included in the study. Although none of these embryos progressed to form blastocysts, they were included in order to determine the pattern of IS activity and SP1 secretory capacity of such 'poor quality embryos' as compared to that of blastocysts.

B.4 IS activity: Materials and Methods

B.4.1 Culture medium samples

Embryos that developed *in vitro* to blastocysts or vacuolated morulae (**VM**) were cultured individually in 1 ml medium (T6 supplemented with 10% autologous maternal serum) in a 24 well Nunclon plate and moved to a well containing fresh medium, from day 3 up to day 14 (see 3.4.7). The medium collected from the previous well was stored at minus 20°C[1]. There was no change in the serum supplementation for individual embryos throughout the entire duration of culture. hCG assay was performed on ECF for a separate study, the results of which have been reported elsewhere (Dokras *et al*, 1991). These ECFs were then refrozen before being assayed at a later date for IS activity. Unfortunately control medium (i.e. T6 plus autologous maternal serum, but no embryo) for each

---

[1] Medium collected from embryos when allotted for research on day 2 contains factors secreted by the embryo over the previous 24 hours since the detection of fertilisation and is therefore referred to as day 1, when referring to embryo-derived products. Similarly medium collected on day 3 of culture is designated day 2 and so on. This holds good for the sense in which 'day' is used when referring to IS activity and SP1 levels in this appendix.

embryo was unavailable for the IS activity study. Since maternal serum is known to cause a certain amount of suppression, it would not be possible to distinguish the levels of embryo-derived suppression from those due to the maternal serum in the medium, without this control. This study therefore makes a comparative assessment of levels of IS activity between the days of culture, as embryo development progressed, rather than measuring absolute levels of embryo-derived IS activity.

B.4.2 Principle of method used to assess IS activity:

The concanavalin A stimulated lymphocyte reaction (Con A SLR) was used to assess the IS activity in the ECF. Proliferation of lymphocytes was stimulated by using Concanavalin A. The effect of the ECF from day 4-14 on the proliferation of lymphocytes was observed and compared with the effect of day 3 ECF. Suppression of lymphocyte proliferation (comparative to day 3) by ECF on any day between day 4-14 would suggest one of two possibilities. Firstly, it could indicate the presence, in the ECF for these days, of suppressor activity due to the active secretion by embryos of some immunosuppressive factor(s). Alternatively, it could suggest the inhibition of the secretion of a stimulatory factor by the embryo thus causing an apparent suppression of immune cell proliferation.

B.4.3 IS Assay: Con A SLR

The steps involved in the Con A SLR were as follows:

Preparation of peripheral blood lymphocytes

Venous blood was drawn from volunteers into 50ml sterile tubes (Nunc Denmark) containing heparin (1000IU/ml) in phosphate buffered saline (PBS). Blood was diluted with an equal volume of PBS, underlayered with 10ml of Ficoll Paque (Pharmacia LKB, Biotechnology, AB Uppsala, Sweden) and then centrifuged for 30 min at 300g. Lymphocytes were aspirated from the interface, washed three times in PBS, resuspended in a dilution of $10x10^6$cells/ml in RPMI (RPMI, 1640 medium with Hepes 25mM and L-glutamine, Life Technologies Ltd., Paisley, Scotland) plus penicillin (100 units/ml) and streptomycin (100μg/ml).

Lymphocyte culture

Lymphocytes were cultured in RPMI supplemented with 10% male AB serum in 96 well round bottomed Nunclon plates (Delta, Nunc, Denmark) at a concentration of $5x10^4$ cells /well in a total volume of 200μl/well. ECF was added in the ratio of 1:5. Depending on the amount of ECF available, it was tested in replicates of at most 5 and, very rarely, at least 2 wells. Concanavalin A at a final concentration of 10μg/ml was used as the mitogenic stimulant. Lymphocytes were cultured for 48 hours at 37°C in a humidified atmosphere of 5% CO_2 in air. In order to quantify levels of lymphocyte proliferation, 0.5μCi/well of methyl tritiated thymidine (Amersham International PLC, Amersham, UK) was added. Cells were harvested on glass fibre filter paper (Wallac Oy, Turku, Finland) 12-18 hours later, using a Skatron Combi cell harvester. Radioactivity was

counted with a liquid scintillation counter (1205, Betaplate TM, Pharmacia.). The uptake of radioactive thymidine was directly proportional to the extent of cell proliferation. All day 3-14 samples for a single embryo were assayed as a batch.

B.4.4 Calculation of IS activity

The median counts per minute (or mean when there were only 2 wells) of each set of replicates was determined and the percentage suppression (or stimulation) calculated as follows

% SUPPRESSION/STIMULATION INDEX =

$$\frac{\text{cpm in day 3 ECF - cpm in day 'n' ECF}}{\text{cpm in day 3 ECF}} \times 100$$

where 'n' is any day from day 4-14

An index above 0 represents immunosuppression and an index below 0 immunostimulation relative to day 3, the first day of culture

B.4.5 IS activity in Fresh versus Frozen ECF

Since the medium used in the Con A SLR assay had previously been frozen, thawed for the hCG assay and then re-frozen, an experiment was conducted to determine whether this could affect the IS activity. The IS activity of fresh culture medium from a further 20 embryos as well as 9 blastocysts cultured individually for 24h, was compared with the activity detected after freeze thawing the medium twice. Minimal levels of

immunosuppressive activity were detected in the fresh samples. This finding persisted even if the concentration of supernatant used in the Con A SLR was doubled or medium in which embryos had been cultured for as long as 5-6 days was used. The levels of IS activity detected in fresh media were not significantly different from the levels detected when the same media were frozen and thawed twice.

B.5 SP1: Materials and Methods

B.5.1 Culture medium samples

Spare embryos were cultured individually in 1 ml medium from day 1 onwards. Those that developed into blastocysts or vacuolated morulae were cultured up to day 14 (refer 3.4.7). Each day they were transferred to fresh medium and the medium collected from the previous well was stored at minus 20°C. In cases where insufficient maternal serum was available, T6 supplemented with 10% pooled serum (from women whose IVF treatment resulted in successful pregnancies) was used. Control medium was available for all embryos.

SP1 levels in the ECF were measured by using an ELISA.

B.5.2 SP1 assay principle

A highly sensitive (10pg/ml) sandwich type double-antibody microplate enzyme immunoassay (ELISA) was used to measure SP1 in the ECF. The rabbit polyclonal anti SP1 IgG was immobilised to the walls of 96 well

plates. The second antibody was coupled to the enzyme conjugate horseradish peroxidase. SP1 was bound between the polyclonal and the enzyme labelled antibody. A quantitative measure was obtained by measuring the absorption in a microplate ELISA reader at 492nm.

The level of embryo-derived SP1 present in the ECF was obtained by subtracting the level (if any) present in the corresponding control medium.

B.5.3 ELISA

High binding capacity 96 well plates (Maxisorp, Nunc, Denmark) were coated the day before the assay with excess rabbit polyclonal anti-SP1 IgG (Dako-immunoglobulins a/s, Denmark) 25-40μg/ml, in 0.05M sodium carbonate, pH 9.6 at 75μl/well. The coating solution was left overnight at 4°C and aspirated next morning. Excess sites were then blocked with 200μl of bovine serum albumin (BSA), 5mg/ml in phosphate buffer (PBS), pH 7.4, for 2h at room temperature. The plates were then washed four times with PBS containing Tween-20 (0.1% v/v, PBS-T). Diluant (high protein carrier buffer BSA, 20mg/ml plus Emulsit, a non-ionic detergent, 0.05% in PBS) was added to the culture medium sample (1:3) and this was then added to the microplate wells in duplicate (75μl/well). After initial incubation at 37°C for 60min, the plate was left overnight at 4°C. Next morning, the wells were washed three times with PBS-T. The enzyme conjugate, horseradish peroxidase coupled to anti-SP1 IgG (Behringwerke, Germany) was diluted 1:100 with the above high protein buffer before use. Incubation was for 60min at 37°C after which the wells were washed four

times with PBS-T and finally with peroxide-free Tween (Pierce, USA). The reaction was developed with 100µl sodium perborate (0.03%) as the substrate and ortho-phenylene diamine (OPD) in citrate-phosphate buffer, 0.1M, pH 5.0 (1mg/ml) as the chromogen, for 30min at room temperature in the dark. The reaction was stopped with 100µl of 2M sulphuric acid, the plates shaken for 1min, and the absorption measured in a Labsystems (Finland) microplate ELISA reader at 492nm. Lyophilised aliquots made from pooled late pregnancy serum, diluted in normal male serum calibrated against the WHO reference preparation 78/610, were used as assay standards. The detection limit of the test was 10pg/ml.

B.6 Note on statistical calculations and tests used

In order to determine the overall rise or fall in levels of IS activity relative to day 3 or to determine the cumulative SP1 secretion for each embryo, the area under the curve obtained by joining points showing the IS activity levels relative to day 3 (or levels of SP1 secretion), for each day was calculated (Matthews *et al*, 1990). The Mann Whitney Test was used to test for differences in IS activity or SP1 secretion between the blastocysts and vacuolated morulae as well as the hatched and intact blastocysts. The Kruskal Wallis test was used to test for differences between the 3 grades of blastocysts. For ease of interpretation, comparison of the cumulative difference in daily IS activity from day 3 (or cumulative levels of SP1 secretion) between different groups has been shown here using notched box plots. The horizontal line separating the notches in the plots

indicates the median while the notches approximate a 95% confidence region around the median. The notched boxes contain 50% of the data, 25% above and 25% below the median. Overlap of the notches indicates the absence of a significant difference between the medians (Axum 4.0 for Windows, 1994).

B.7 IS activity: Results and discussion

B.7.1 Results

IS activity in the culture medium from 16 blastocysts (5 **BG1**, 5 **BG2**, 6 **BG3**) and four **VM** was measured daily from day 3 to day 14. Levels of IS activity in the ECF of each embryo for each day were compared with the levels for that embryo detected in the day 3 ECF.

There was no consistent pattern to the IS activity over the culture period in the embryos belonging to any of the four groups investigated. Table B.1 shows the levels of IS activity relative to day 3 for the three grades of blastocysts and the **VM.**

Table B.1 Level of IS activity (mean ± standard error of mean)[2] relative to day 3

Day	BG1	BG2	BG3	VM
4	-184.96±133.63	22.14±M	-26.09±77.22	0.00±20.84
5	-30.47±42.44	4.34±25.13	43.74±27.66	4.26±8.39
6	-21.41±19.00	26.14±11.55	-91.83±128.84	15.59±17.92
7	-6.53±12.89	8.31±16.24	-18.36±40.54	12.34±5.72
8	-124.92±91.32	1.22±28.80	0.93±54.55	-5.92±16.42
9	-30.23±57.55	0.48±19.96	-53.93±109.88	4.17±15.02
10	-68.27±58.10	31.94±18.06	17.78±36.49	9.66±13.94
11	-29.92±47.88	-27.95±42.87	-104.96±110.55	-10.46±37.07
12	-19.68±37.07	16.20±5.65	-13.29±40.06	26.59±20.85
13	-17.26±4.42	1.37±12.68	-147.52±124.92	34.65±14.37
14	-27.30±22.60	7.42±38.41	-45.10±66.08	30.58±M

Figure B.1 shows the pattern of IS activity relative to day 3 for the 16 blastocysts and four VM. The amount that the IS activity over days 4 - 14 deviated from that on day 3 for blastocysts was not significantly different to that for the vacuolated morulae (Figure B.2). Similarly the three grades of blastocysts could not be distinguished by their pattern (Figure B.3) of activity as well as showed no significant difference between the groups (Figure B.4). In addition, the 'good' quality BG1's or BG2's did not show any particular rise in levels of IS activity around days 5 - 7 (i.e. around the time of implantation) compared to the earlier days of culture. Further, the hatched and intact blastocysts showed no discriminatory

[2] Mean values on any day calculated by averaging over embryos for which ECF was available for that day

pattern (Figure Figure B.5) and no significant difference in the amount that daily IS activity differed from that on day 3 (Figure B.6).

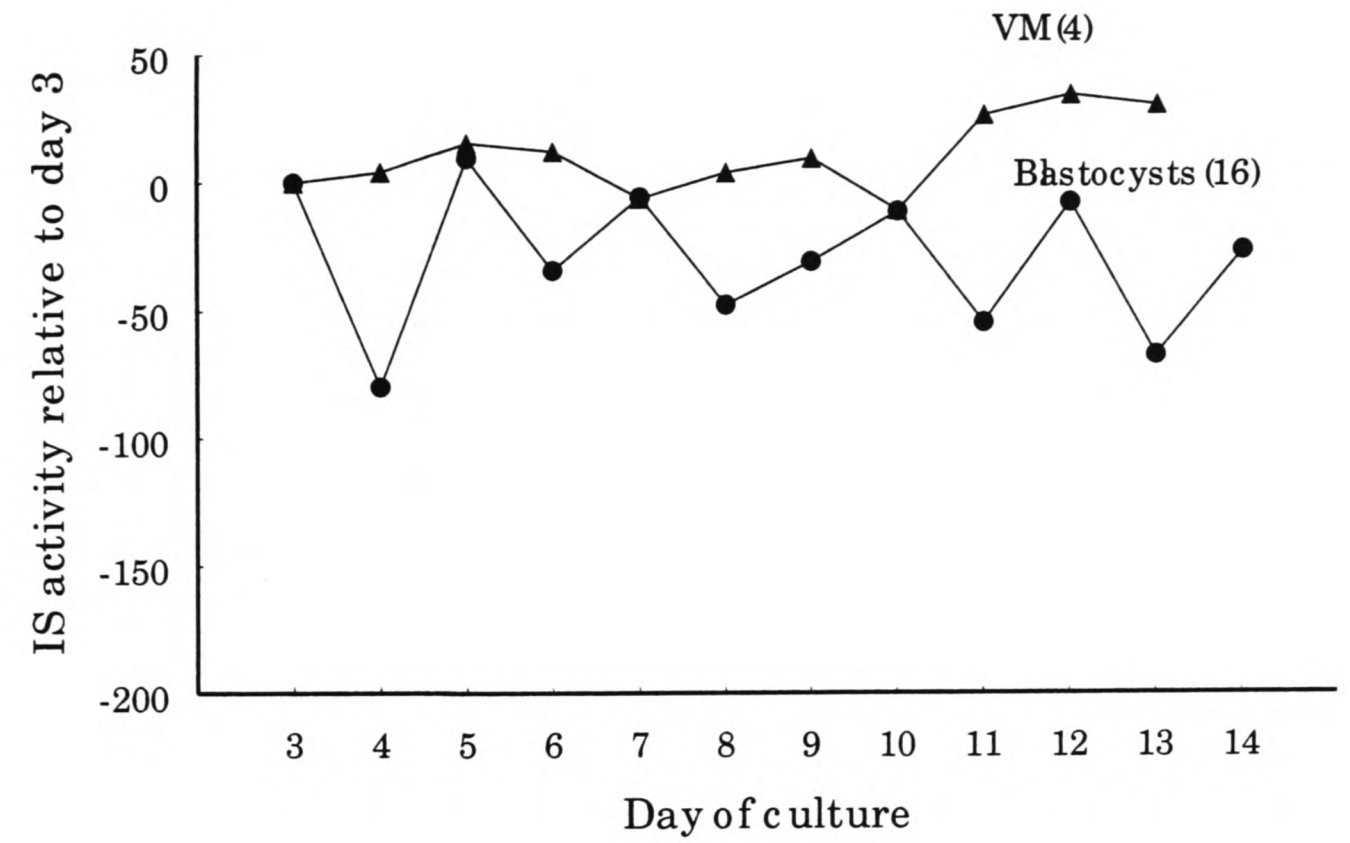


Figure B.1 Mean daily IS activity relative to day 3 for 16 blastocysts and 4 VM - day 4 to 14

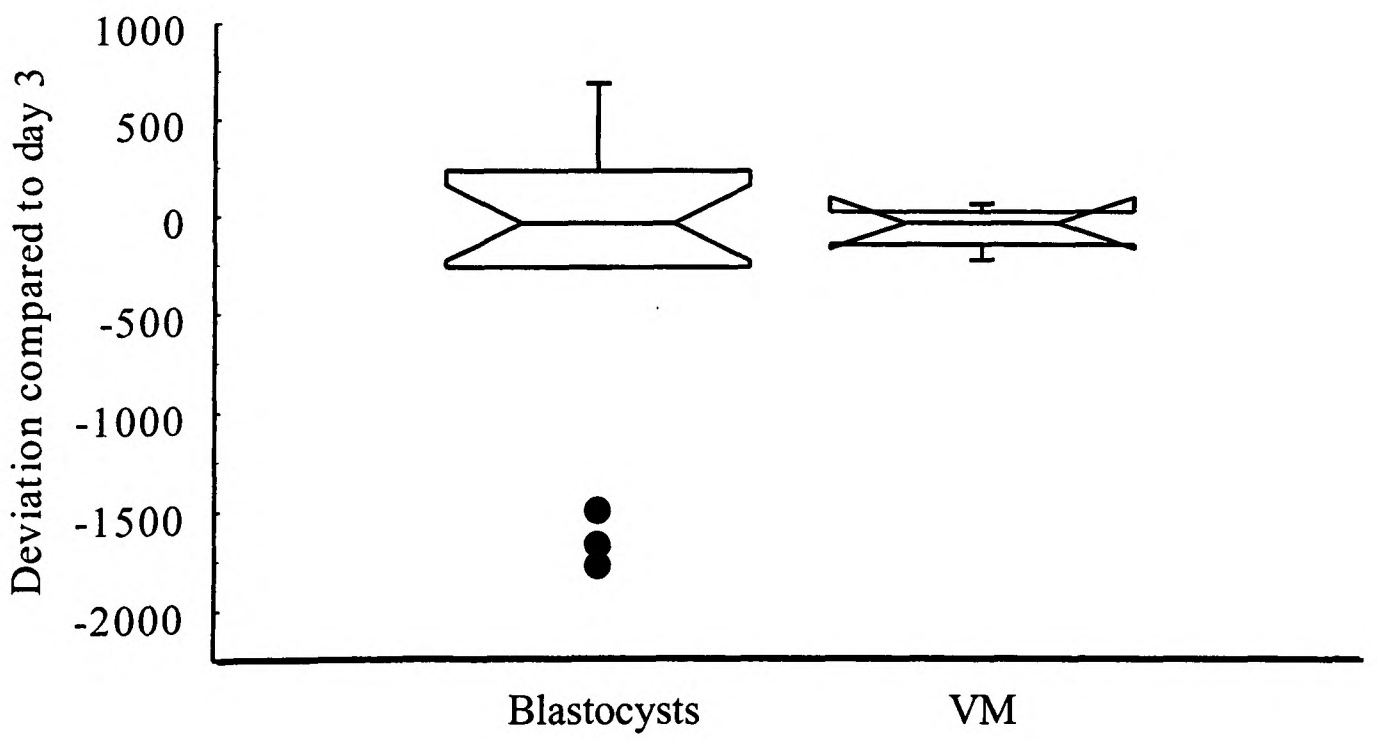


Figure B.2 Statistical comparison of cumulative deviation of IS activity from that on day 3 for 16 blastocysts and 4 VM

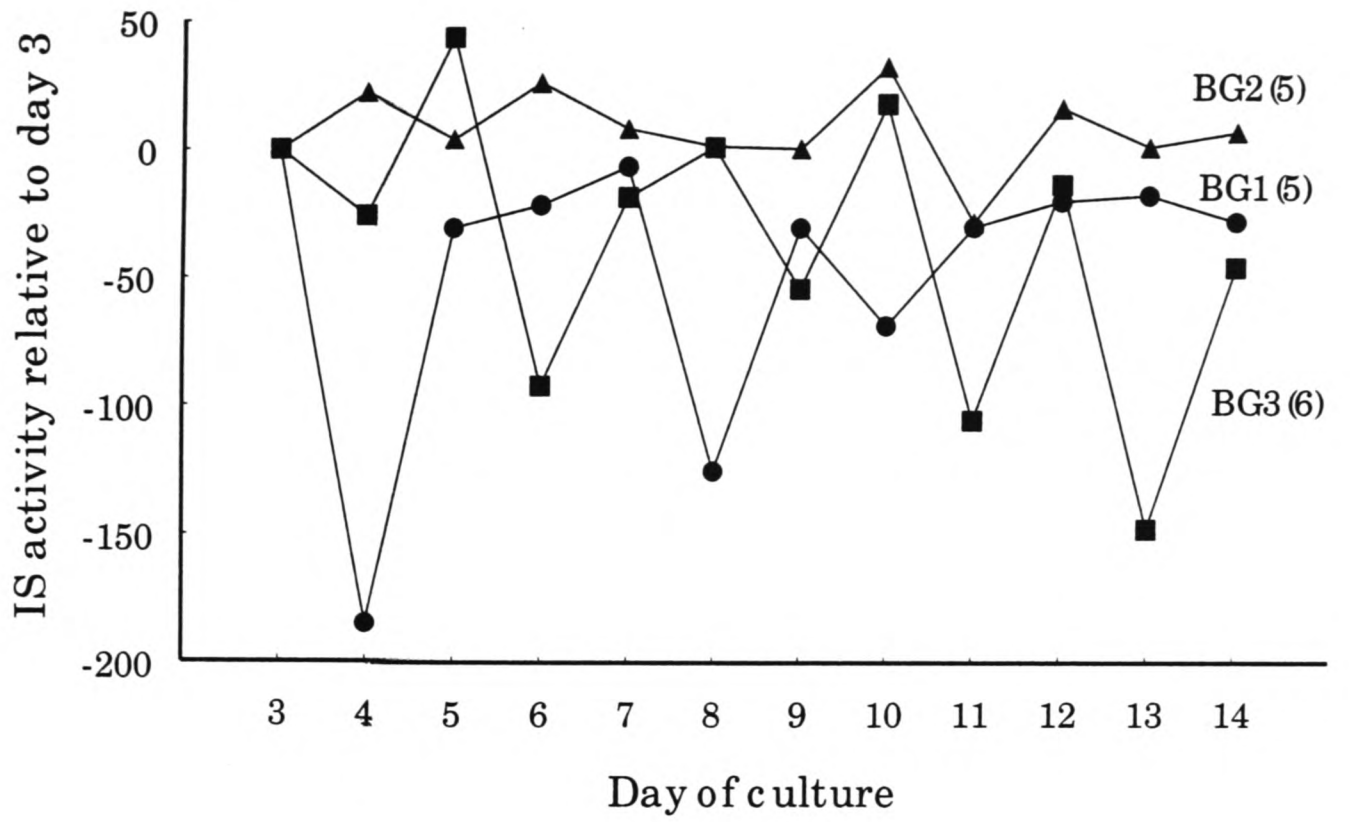


Figure B.3 Mean daily IS activity relative to day 3 for 5 BG1, 5 BG2 and 6 BG3 - day 4 to 14

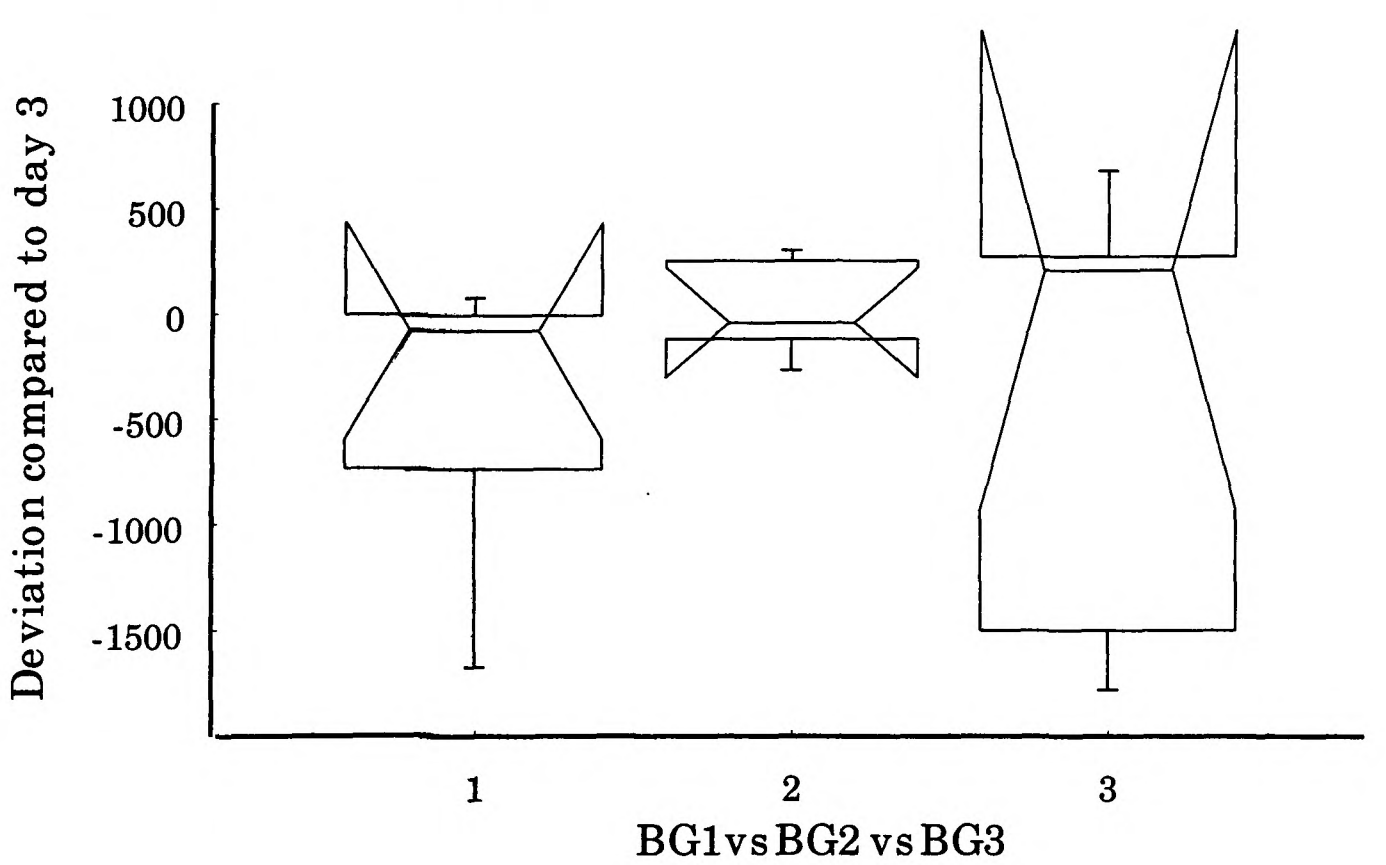


Figure B.4 Statistical comparison of cumulative deviation of IS activity from that on day 3 for 5 BG1, 5 BG2 and 6 BG3

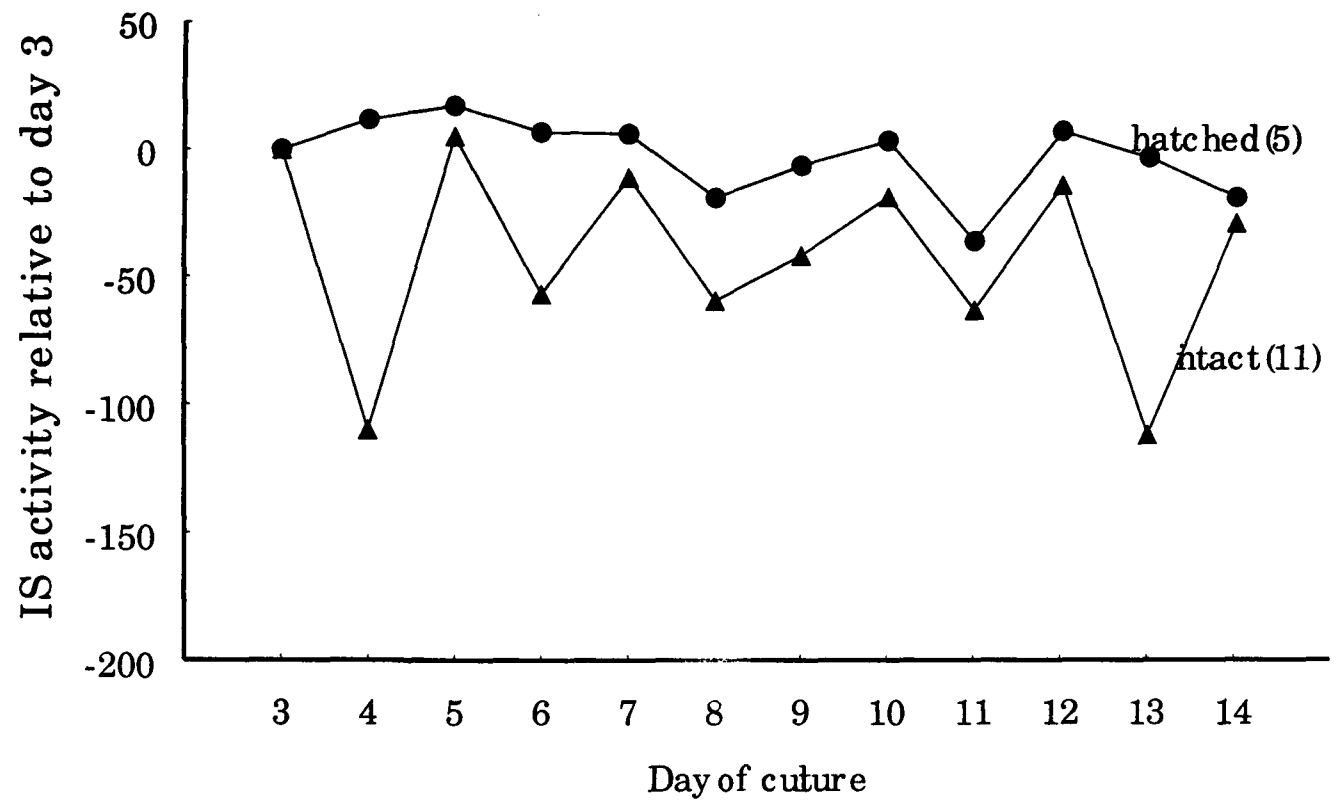


Figure B.5 Mean daily IS activity relative to day 3 for 5 hatched and 11 intact blastocysts - day 4 to 14

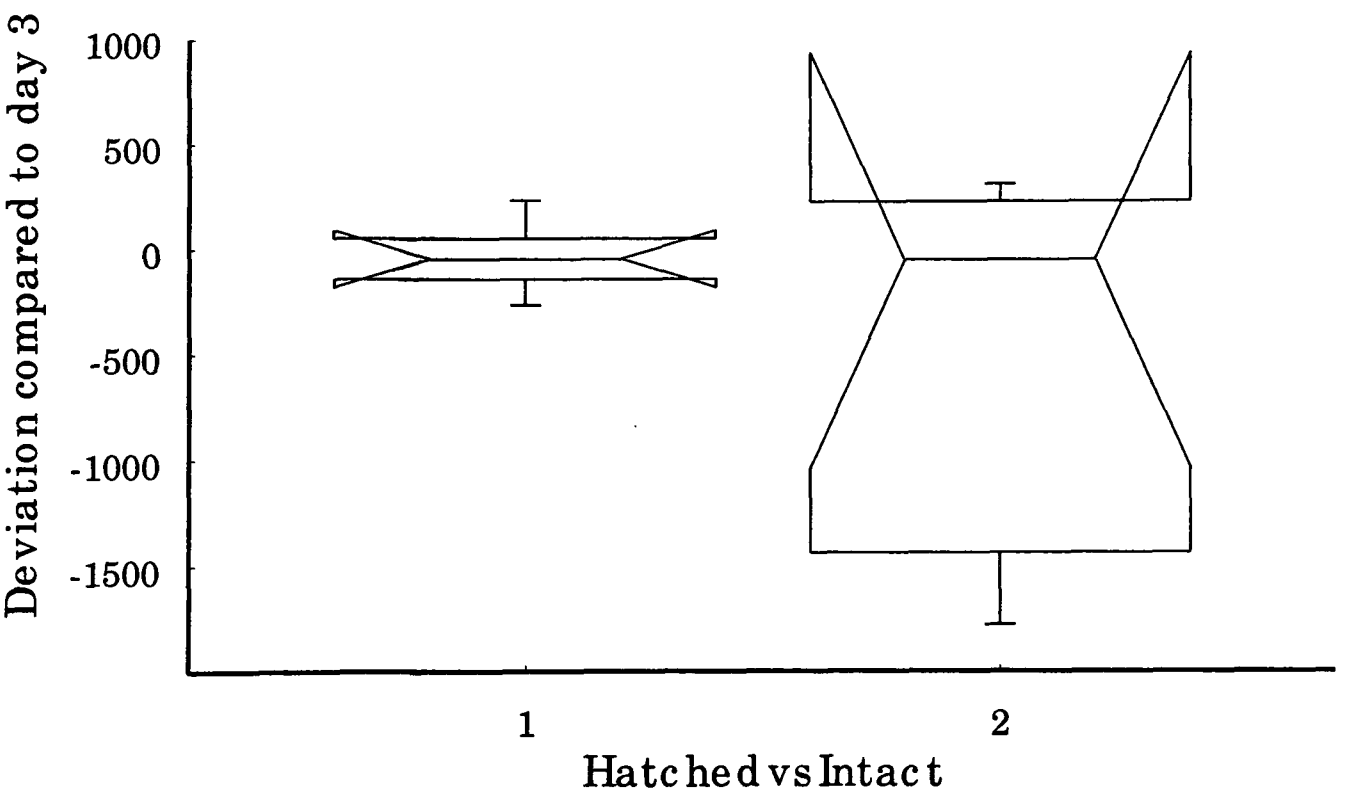


Figure B.6 Statistical comparison of cumulative deviation of IS activity from that on day 3 for 5 hatched and 11 intact blastocysts

B.7.2 Discussion

Most studies to date have investigated the correlation between the *levels* of IS activity in ECF and the outcome of the embryo transfer. In contrast to previous studies, this study investigates the *pattern* (rather than the level) of IS activity in ECF from the early embryo right through to the time around and beyond implantation.

The IS activity in the ECF did not show any definite pattern over the 11 days for embryos belonging to any of the four groups. Neither the blastocysts nor the VM showed any tendency towards higher levels around implantation. The three grades of blastocysts did not demonstrate any definite patterns of secretion that could distinguish them from each other. Furthermore, blastocysts that hatched did not show any difference in patterns from the intact blastocysts.

The findings indicate that there is no association between the morphological grading (and therefore blastocyst quality as suggested by the grading) and the IS activity pattern of embryos. Measuring levels of IS activity in ECF, Armstrong and colleagues too demonstrated no significant difference between embryos of differing quality, i.e. between cleaving or non-cleaving ova and pregnant or non-pregnant groups. The authors propose mechanisms other than embryo derived IS factor(s) that could play a role in protecting the semi-allogenic embryo from immune rejection (Armstrong *et al*, 1989 and Armstrong and Chaouat, 1993).

B.8 SP1 Secretion: Results and Discussion

B.8.1 Results

SP1 secretion by 32 blastocysts and six vacuolated morulae was measured daily from day 1 to day 14. Levels of SP1 secreted by each embryo over each day were obtained by subtracting the levels detected in the respective culture medium control from the levels detected in the ECF. Of the blastocysts, SP1 secretion was detected in 17/19 **BG1**, 4/4 **BG2** and 6/9 **BG3**. 4/6 **VM** secreted SP1. The mean daily SP1 secretion for all blastocysts and **VM** is shown in Figure B.4. The earliest detection of SP1 (i.e. levels > 10pg/ml) was on day 1 in 8 out of the 38 embryos. Of these, two were **BG1**, two **BG2**, one **BG3** and three **VM**. No individual embryo secreted SP1 over all **14** days of culture and the maximum number of days over which SP1 was detected was 10 for one **BG1**. Five blastocysts and one **VM** showed detectable secretion over one day only. Five out of the 38 embryos continued to secrete SP1 on day 14. Of these, three were **BG1** and one each of **BG2** and **BG3.**

The 32 blastocysts showed significantly higher (p = 0.02) cumulative SP1 secretion as compared to the 6 **VM** (Figure B.8 and Figure B.9). All three grades of blastocysts also showed a consistent peak in the mean level of secretion on day 3 that was absent in the vacuolated morulae (Table B.2 and Figure B.7).

Table B.2 Daily SP1 secretion (mean ± standard error of mean)[3] in pg/ml from day 1 to day 14

Day	BG1	BG2	BG3	VM
1	5.93±3.25	4.25±4.25	7.56±6.40	8.80±5.43
2	21±10.61	45.67±45.67	19.22±11.56	1.67±1.67
3	38.50±21.76	71.50±60.70	23.44±16.26	2.83±2.83
4	6.59±3.28	5.25±5.25	3.89±3.89	0.00±0.00
5	7.72±3.80	0.00±0.00	9.88±6.55	0.00±0.00
6	8.68±4.87	58.00±58.00	6.44±6.44	1.67±1.67
7	22.58±7.67	7.25±7.25	2.00±2.00	5.83±4.17
8	13.44±6.58	0.00±0.00	2.22±2.22	0.00±0.00
9	22.11±9.83	7.25±7.25	2.00±2.00	0.00±0.00
10	25.16±8.38	81.50±72.78	13.89±8.35	4.00±4.00
11	40.44±16.46	58.00±58.00	2.11±2.11	0.00±0.00
12	45.22±21.87	42.00±42.00	10.88±7.17	3.83±3.83
13	36.18±21.10	7.50±7.50	13.00±13.00	0.00±0.00
14	33.35±23.40	7.75±7.75	7.75±7.75	0.00±0.00

BG1 and **BG2** show a trend towards higher daily cumulative SP1 levels than **BG3** (Figure B.10). On statistical analysis, however, the three grades of blastocysts showed no significant difference in either the daily or

[3] Mean values on any day calculated by averaging over embryos for which ECF was available for that day.

cumulative SP1 levels (Figure B.11). When the 12 blastocysts that hatched were compared with those that remained intact within the zona, there was no significant difference in the cumulative secretion (Figure B.12). For individual embryos, there was no correlation between the level of secretion and the morphology and development of the embryo.

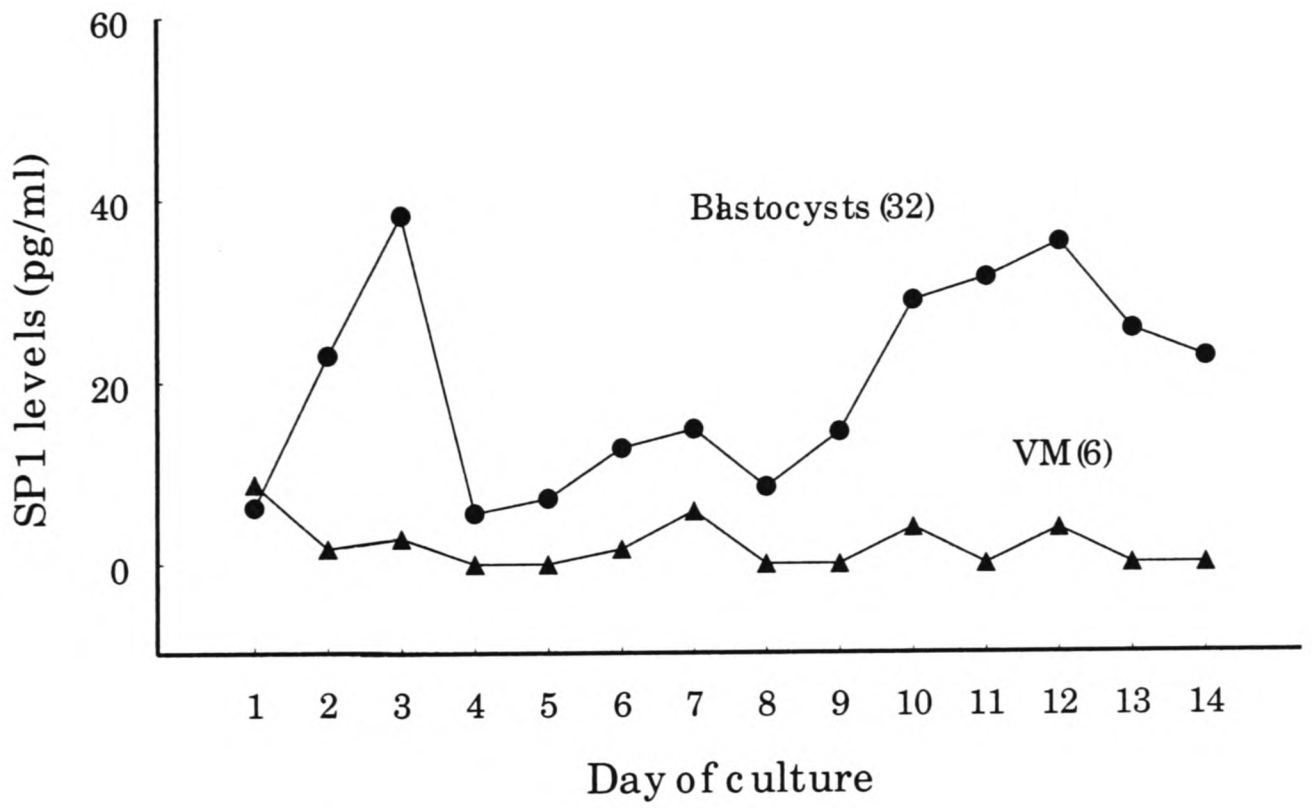


Figure B.7 Mean daily SP1 secretion by 32 blastocysts and 6 VM - day 1 to 14

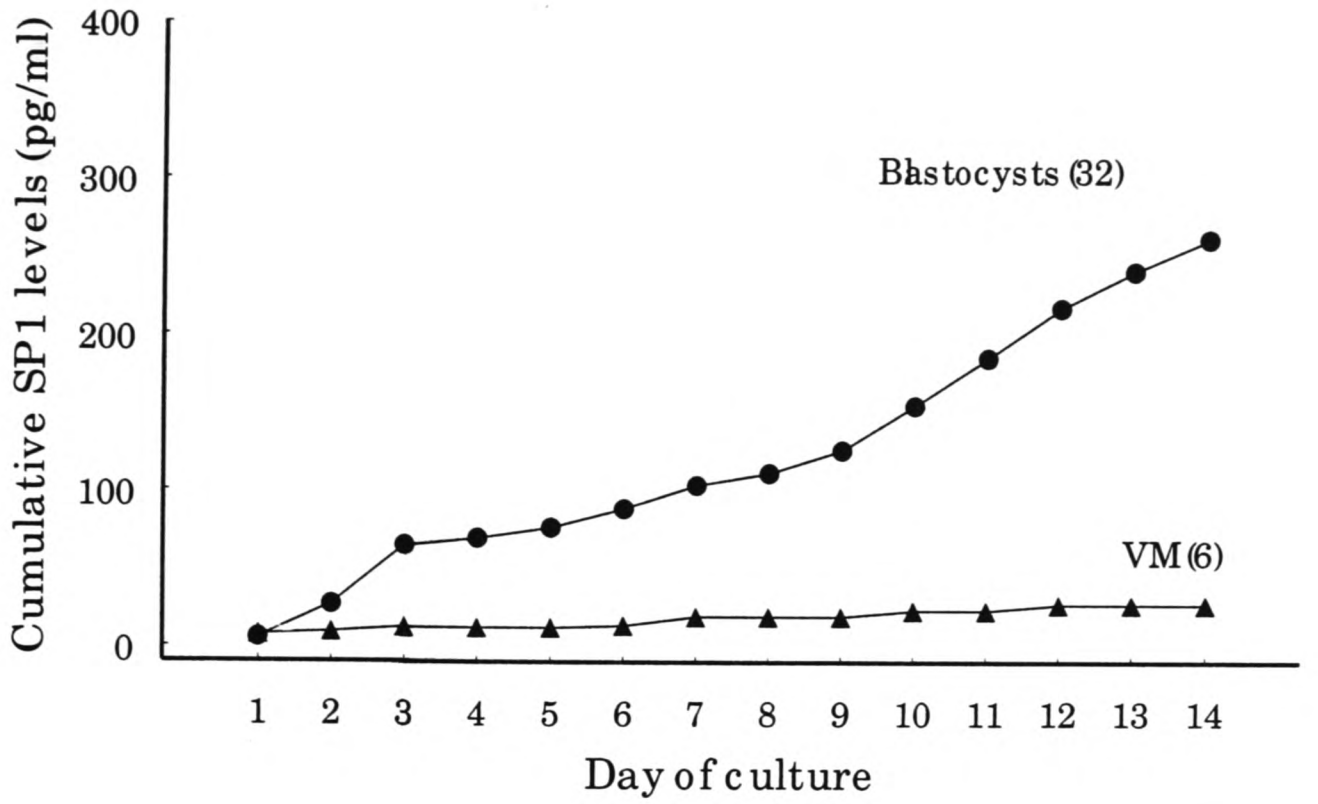


Figure B.8 Mean cumulative SP1 levels of 32 blastocysts and 6 VM - day 1 to 14

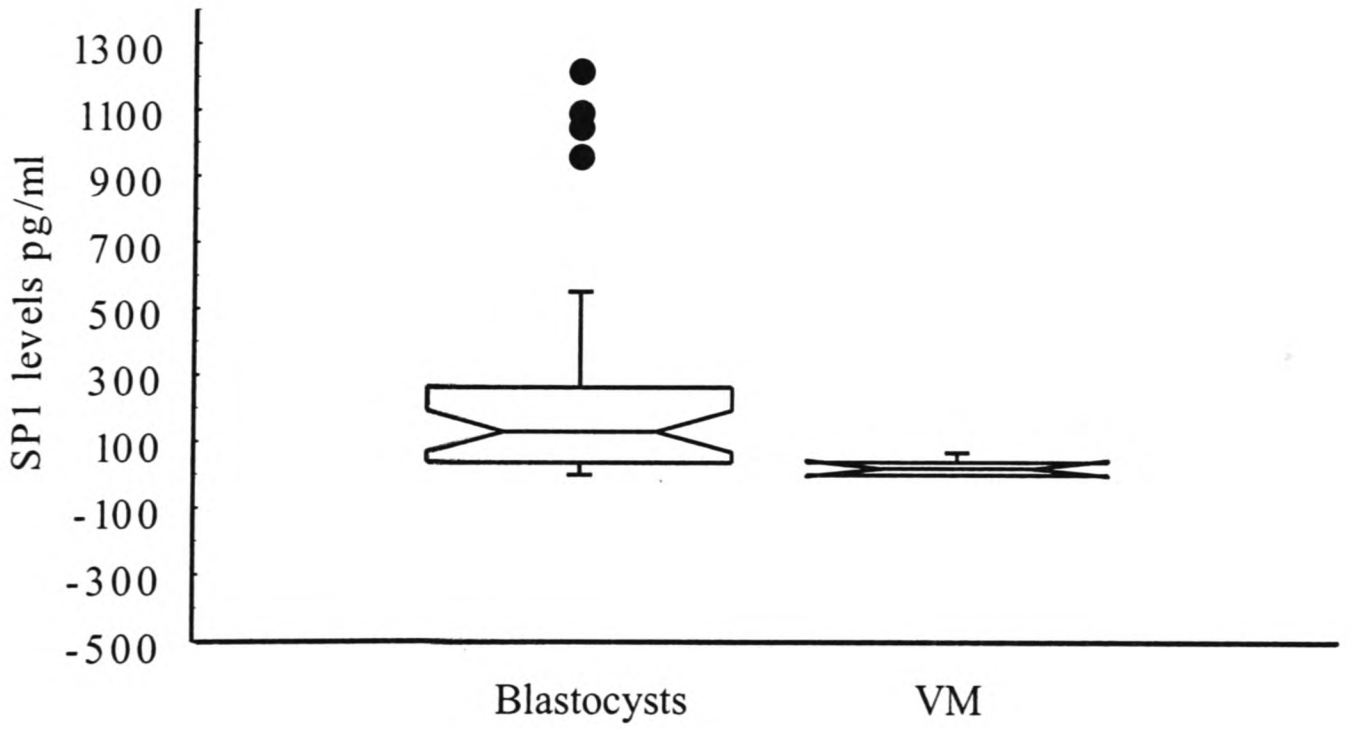


Figure B.9 Statistical comparison of cumulative SP1 secretion of 32 blastocysts and 6 VM

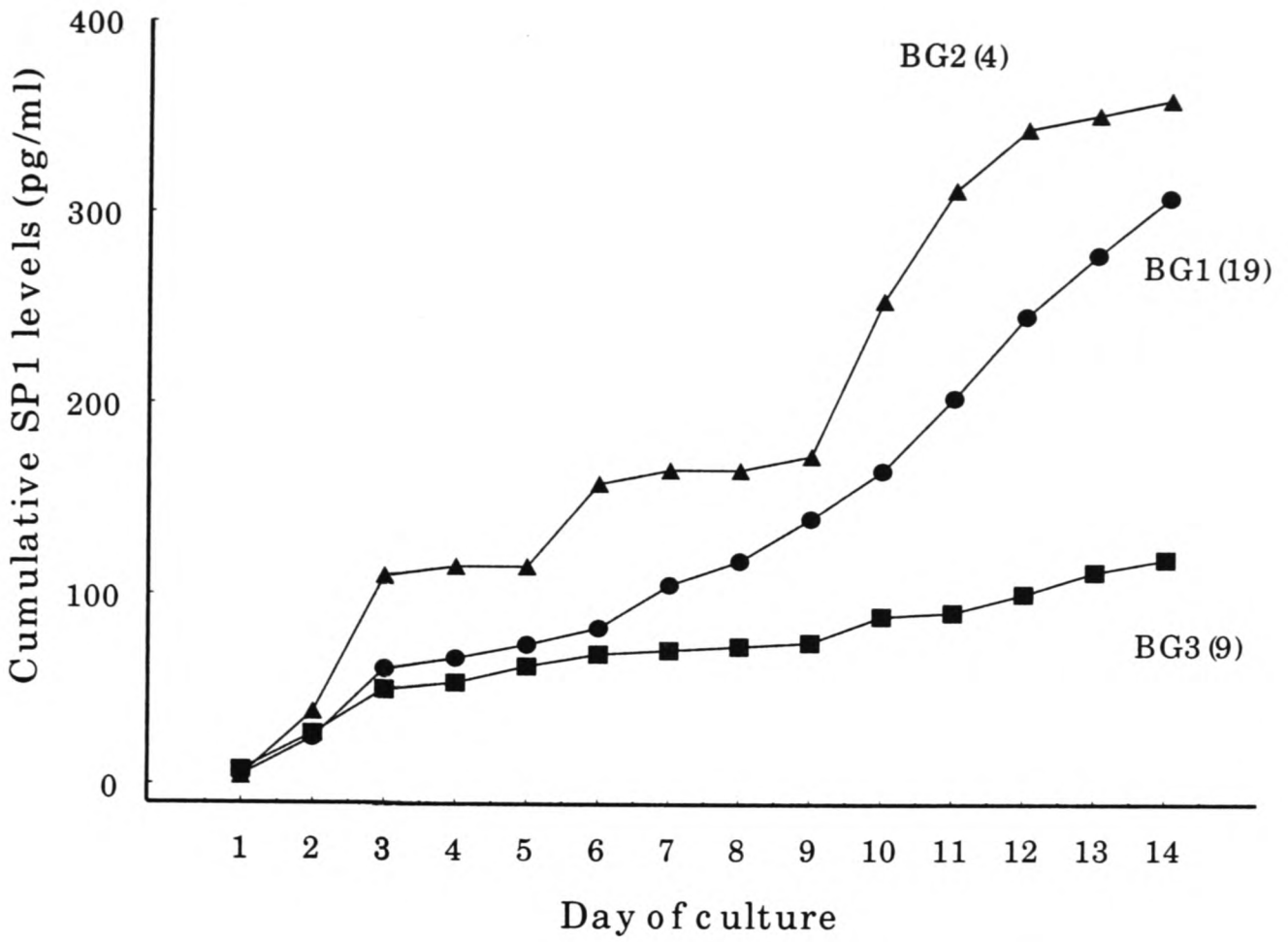


Figure B.10 Mean cumulative levels of 19 BG1, 4 BG2 and 9 BG3 - day 1 to 14

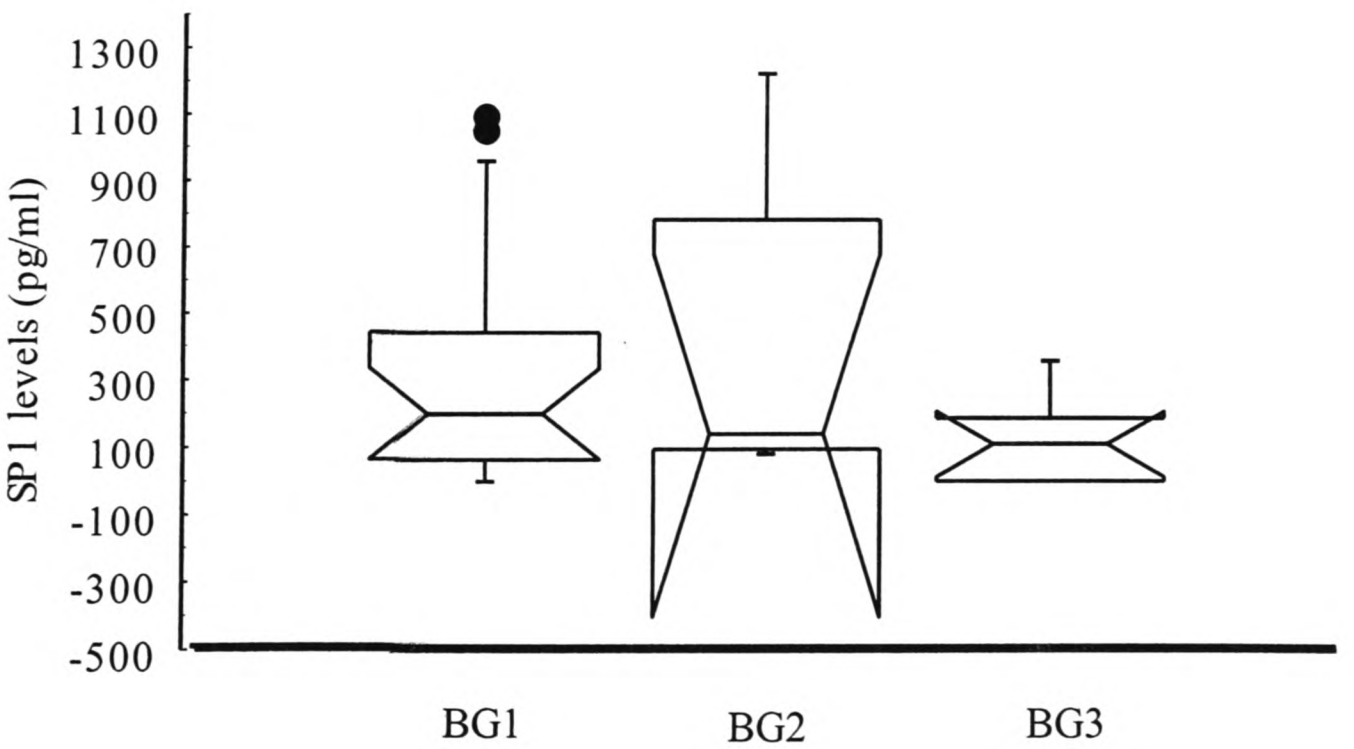


Figure B.11 Statistical comparison of cumulative SP1 secretion by 19 BG1, 4 BG2 and 9 BG3

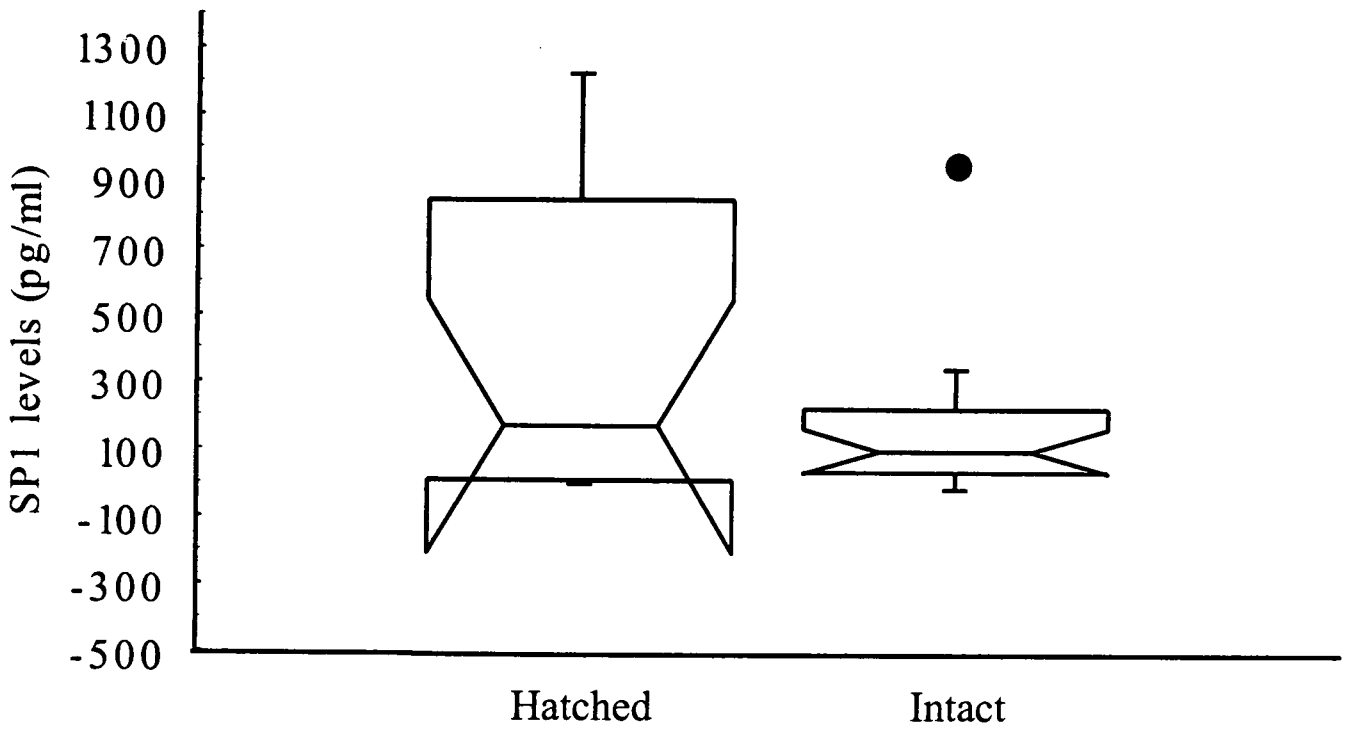


Figure B.12 Comparison of cumulative SP1 secretion by 12 hatched and 20 intact blastocysts

B.8.2 Discussion

SP1 levels secreted by embryos, classified on the basis of their morphology, were measured to investigate the association between the morphological grade of the blastocyst and the levels of SP1 secreted. Using a highly sensitive immunoassay, low levels of SP1 were detected in the ECF confirming earlier findings (Dimitriadou *et al*, 1992). While Dimitriadou *et al* studied the levels of SP1 in ECF from day 3 onwards, here SP1 secretion in the ECF could be detected as early as day 1. SP1 levels secreted by the blastocysts were significantly higher than those by the vacuolated morulae suggesting that in general there is a correlation between SP1 and embryo quality. In addition embryos developing into blastocysts showed a pattern of secretion that differed from the **VM** with

all 3 grades showing a distinct peak of secretion on day 3. This peak is of particular interest as it could have the potential to discriminate between embryos destined to develop into blastocysts and those destined to form vacuolated morulae. However, on analysis, neither the level of SP1 secretion on day 3 nor the cumulative day 1 to day 3 levels showed a statistical difference.

Blastocysts of differing developmental potential, as indicated by their ability or inability to hatch, could not be distinguished by the levels of SP1 secretion. Although the better grade (as defined by morphology) blastocysts **BG1** and **BG2** showed a trend towards higher daily mean cumulative SP1 levels than **BG3,** there was no significant difference in the cumulative or daily SP1 secretion between the 3 grades of blastocysts. Despite the initial promise shown by differences in SP1 levels, between embryos of differing quality, (i.e. blastocysts versus vacuolated morulae), the lack of a significant difference in SP1 levels secreted by blastocysts belonging to the three morphological grades suggests that there is no correlation between the morphological grading (and therefore blastocyst quality as suggested by the grading) of the blastocyst and its ability to secrete SP1. Although a correlation between levels of SP1 in maternal serum and foetal well-being has been reported, it is possible that SP1 secretion is not sufficiently defined at the earlier stages of embryo development to demonstrate such correlations. This is supported by the finding here that SP1 is neither secreted by all blastocysts nor was it detected for individual blastocysts on all days of culture. Furthermore,

the first day of starting secretion was highly variable ranging from day 1 for six blastocysts to as late as day 7 for three, and days 8 and 9 for one each respectively. Once secretion was detected it did not necessarily persist but stopped and restarted at irregular intervals.

The levels of SP1 detected here were much lower than those reported by Pope *et al* (1984) for baboon embryos and Dimitriadou *et al* (1992) for human embryos. Possible causes for this could be the assay used or sub-optimal culture conditions. The highest level of SP1 detected in our study was 391pg/ml/day secreted by a grade 1 blastocyst on day 3 of culture in contrast to 4800pg/ml/3days reported by Dimitriadou *et al* (1992). This could have been due to the difference in culture media (they used Ham's F-10 supplemented with 7% maternal serum) which can affect embryo quality and thus secretory capacity. However this is unlikely as blastocysts of an obviously good quality, as indicated by their hatching, showed levels of SP1 secretion much lower than those reported by Dimitriadou *et al* (1992). Another explanation could be the that in this study the medium was changed daily whereas Dimitriadou *et al* changed the medium only once in 3 days for most embryos. It is possible that frequent changes of the culture medium could deplete embryo-derived factors that stimulate SP1 secretion. Unfortunately no direct comparison of SP1 levels in culture medium using the assay system employed by Dimitriadou *et al* is possible as the Enzygnost SP1, Behringwerke kit used by them is no longer commercially available.

B.9 Conclusions

The pattern of IS activity in ECF as well as the levels of SP1 secretion by embryos do not appear to show any association with the morphological grading scheme of Dokras *et al* (1993). In view of the small number of blastocysts investigated and the high variance in the levels of SP1 as well as the IS activity however, no firm conclusions can be drawn from these experiments.

B.10 Summary

IS activity and SP1 secretion in ECF were assessed up to day 14 using a Con A SLR and an ELISA. Their association with a morphological blastocyst grading scheme was investigated and in the data available here no correlation was identified.

References.

References

Acosta AA, Chillik K, Brugo S, Swanson RJ, Pilleban P, Yan J Haque D: In vitro fertilization and the male factor. Urology 28:1-7, 1986.

Ahmed AG, Klopper A: Determination of the stage of gestation by the assay of chorionic gonadotrophin and Schwangerschaftsprotein 1. Br J Obstet Gynaecol 91:1234-1239, 1984.

Andersen CY: Concentrations of free oestradiol and progesterone in human preovulatory follicular fluid. Hum Reprod 6:359-364, 1991

Andersen CY: Levels of steroid binding proteins and steroids in human preovulatory follicle fluid and serum as predictors of success in in vitro fertilization embryo transfer treatment. J Clin Endocrinol Metab 71:1375-1381, 1990.

Armstrong DT, Chaouat G, Guichard A, Cedard L, Andreu G, Denver L: Lack of correlation of immunosuppressive activity secreted by human in vitro fertilized (IVF) ova with successful pregnancy. J In Vitro Fert Embryo Transf 6:15-21, 1989.

Armstrong DT, Chaouat G: Immunological aspects of human in vitro fertilization and embryo transfer. In: Immunology of Pregnancy, CRC Press, Boca Rato:33-46,1993.

Axum 4.0 for Windows: Box Plots in Technical graphics and data analysis. Trimetrix, 1994.

Basuray R, Rawlins RG, Radwanska E, Henig I, Sachdeva S, Tummon I, Binor Z, Dmowski WP: High progesterone/estradiol ratio in follicular fluid at oocyte aspiration for in vitro fertilization as a predictor of possible pregnancy. Fertil Steril 49:1007-1011, 1988.

Biggers JD, Bell JE, Benos DJ: Mammalian blastocyst: transport functions in a developing epithelium. Am J Physiology 255:C419-C432, 1988.

Bischof P: Placental Proteins. Contributions to Obstetrics and Gynecology 12: 6-22, 1984.

Bolton VN, Hawes SM, Taylor CT, Parsons JH: Development of spare human pre-implantation embryos in vitro: an analysis of the correlations among gross morphology, cleavage rate and development to blastocyst. J in vitro fertil embryo transfer 6:30-35, 1989.

Bongso A, Fong CY, Ng SC, Ratnam S: Human embryonic behaviour in a sequential human oviduct-endometrial coculture system. Fertil Steril 61:976-978, 1994.

Bose R, Mahadevan MM: Immunosuppressive activity in human embryo growth media is associated with successful pregnancy effect of gonadotropin releasing hormone agonist (GnRHa) treatment of patients undergoing in vitro fertilization and embryo transfer (IVF ET). J Clin Immunol 10:175-181, 1990.

Bose R: Human embryo associated immunosuppressor factor(s) from pre and post implantation stages share some similarities. Immunol Letters 20:261-267, 1989.

Bose R: The partially purified pre-implantation suppressor factor may be one of several factors to play a role in successful pregnancy. Immunol Lett 27:119-125, 1991.

Bratko I; Mozetic I, Lavrac N.: Kardio: a study in deep and qualitative knowledge for expert systems. MIT Press, Cambridge, USA, 1989..

Braude P, Bolton V, Moore S: Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature 332:459-461, 1988.

Brison DR, Leese HJ: Role of chloride transport in the development of the rat blastocyst. Biol Reprod 48:692-702, 1993.

Carson RS, Trounson AO, Findlay JK: Successful fertilisation of human oocytes in vitro: concentration of estradiol 17 beta, progesterone and androstenedione in the antral fluid of donor follicles. J Clin Endocrinol Metab 55:798-800, 1982.

Chabab A, Hedon B, Arnal F, Diafouka F, Bressot N, Flandre O, Cristol P: Follicular steroids in relation to oocyte development and human ovarian stimulation protocols. Hum Reprod 1:449-454, 1986.

Ciriminna R, Schillaci R, Cefalu E, Agrifoglio V, Cittadini E: Vero cell effect on "in vitro" human embryos development J Assist Reprod genet 12:134S Abstract PP-46, 1995.

Claman P, Armant DR, Seibel MM, Wang T, Oskowitz SP, Taymor ML: The impact of embryo quality and quantity on implantation and the establishment of viable pregnancies. J In Vitro Fert Embryo Transf 4:218-222, 1987.

Clark DA, Lee S, Fishell S, M[illegible]dev[illegible] [illegible]all H, Mo[illegible]e MA, Schechter O, Stedronska-c[illegible]

J: Immunosuppressive a[illegible]man in vitro fertilization (IVF) culture supernatants and prediction of the outcome of embryo transfer : a multicenter trial. J In Vitro Fert Embryo Transfer 6:51-58, 1989.

Clark P, Niblett T: The CN2 algorithim. Machine Learning 3:261-283, 1989.

Cohen J, Inge KL, Suzman[illegible] Wright G: Videocinematograph[illegible] of fresh and cryopreserved [illegible]sis of [illegible]ic morphology and implantation. Fertil Steril 51:820-827, 19[illegible].

Cohen J, Simons RS, Fehilly CB, Edwards RG: Factors affecting survival and implantation of cryopreserved human embryos. J In Vitro Fert Embryo Transf 3 (1):46-52, 1986.

Collier M, O"Neill C, Ammit AJ, Saunders DM: Measurement of human embryo-derived platelet-activating factor (PAF) using a quantitative bioassay of platelet aggregation. Hum Reprod 5:323-328, 1990.

Conaghan J, Hardy K, Handyside AH, Winston RML, Leese HJ: Selection criteria for human embryo transfer: a comparison of pyruvate uptake and morphology. Jornal of Assisted Reproduction and Genetics 10:21-30, 1993.

Conaghan J, Martin K, Hardy K, Winston NJ, Dawson KJ, Winston RML, Leese HJ: Can non-invasive measurement of pyruvate uptake in culture assist embryo selection for transfer following in vitro fertilization. J Obstet Gynaecol 11:87-88, 1991.

Confino E, Demir RH, Friberg J, Gleicher N: The predictive value of hCG subunit levels in pregnancies achieved by in vitro fertilization and embryo transfer: an international collaborative study. Fertil Steril 45:526-531, 1986.

Cornwallis CM, Skinner SL, Nayudu PL, Lopata A, Thatcher RL, Yeung SP, Whitworth JA : Follicular fluid renin concentration and IVF outcome. Hum Reprod 5:413-417, 1990.

Cummins JM, Breen TM, Harrison KL, Shaw JM, Wilson LM, Hennessey JF: A formula for scoring human embryo growth rates in in vitro fertilization: its value in predicting pregnancy and in comparison with visual estimates of embryo quality. J In Vitro Fert Embryo Transf 3:284-295, 1986.

Dawson K, Margara RA, Hillier SG, Packham D, Winston N, Winston RML: Factors affecting success at the time of embryo transfer. Hum Reprod, Abstracts from the third meeting of ESHRE, abst 134:39, 1987.

Daya C, Clark DA: Production of immunosuppressor factor(s) by preimplantation human embryos. Am J Reprod Immunol. 11. 98-101, 1986.

Daya S, Lee S, Underwood J, Mowbray J, Craft I, Clark D: Prediction of outcome following transfer of in vitro fertilized human embryos by measurement of embryo-associated suppressor factor: In: Clark DA, Croy BA (eds) Reproductive Immunology, Elsevier Science Publishers BV (Biomedical Division) 277-285, 1986.

De Geyter C, De Geyter M, Schneider HP, Nieschlag E: Subnormal sperm parameters in conventional semen analysis are associated with discrepancies between fertilization and pregnancy rates in in vitro fertilization and embryo transfer. Int J Androl 15:485-497, 1992.

Dey SK, Stechschulte DJ, Abdou NI: Modulation of the in vitro lymphocyte response by rabit blastocysts. J Reprod Immunol 3:141, 1981.

Dimitriadou A, Phocas I, Mantzavinos T, Sarandakou A, Rizos D Zourlas PI: Discordant secretion of pregnancy specific β1-glycoprotein and human chorionic gonadotrophin by human pre-embryos cultured in vitro. Fertil Steril 57:631-636, 1992.

Dokras A: Human Trophectoderm biopsy and its application. D.Phil thesis, University of Oxford, Ch 1 and 2:pp19-76, 1992.

Dokras A, Sargent IL, Barlow DH: Human blastocyst grading: an indicator of developmental potential? Hum Reprod 8:2119-2127, 1993.

Dokras A, Sargent IL, Ross C, Gardner RL and Barlow DH: Trophectoderm biopsy in human blastocysts. Hum Reprod 5:821-825, 1990.

Dokras A, Sargent IL, Ross C, Gardner RL, Barlow DH: The human blastocyst: morphology and human chorionic gonadotrophin secretion in vitro. Hum Reprod 6:1143-1151, 1991.

Dor J, Rudak E, Mashiach S, Nebel L, Serr DM, Goldman B: Periovulatory 17 beta-estradiol changes and embryo morphologic features in conception and nonconceptional cycles after human in vitro fertilization. Fertil Steril 45:63-68,1986.

Edirisinghe WR, Murch AR, Yovich JL: Cytogenetic analysis of human oocytes and embryos in an in vitro fertilization programme. Hum Reprod 7:230-236, 1992.

Edwards RG, Bavister BD, Steptoe PC: Early stages of fertilisation in vitro of human oocytes matured in vitro. Nature 221:632-635, 1969.

Edwards RG, Fischel SB, Cohen J, Fehilly CB, Purdy JM, Slater JM, Steptoe PC, Webster JM: Factors influencing the success of in vitro

fertilization for alleviating human infertility. J In Vitro Fert Embryo Transf 1:3-23, 1984.

Edwards RG, Steptoe PC, Purdy JM: Establishing full-term human pregnancies using cleaving embryos grown in vitro. Br J Obstet Gynaecol 87:737-756, 1980.

Enginsu ME, Pieters MH, Dumoulin JC, Evers JL, Geraedts JP: Male factor as determinant of in vitro fertilization outcome. Hum Reprod 7:1136-1140, 1992.

Erenus M, Zouves C, Rajamahendran P,Leung S, Fluker M, Gomel V: The effect of embryo quality on subsequent pregnancy rates after in vitro fertilization. Fertil Steril 56:707-710, 1991.

Everitt BS: The analysis of contingency tables, 2nd ed., Chapman and Hall, London, 1992.

Fehilly CB, Cohen J, Simons RF, Fishel SB, Edwards RG: Cryopreservation of cleaving embryos and expanded blastocysts in the human : a comparitive study. Fertil Steril 44:638-644, 1985.

Feldberg D, Farhi J, Dicker D, Ashkenazi J, Shelef M, Goldman JA: The impact of embryo quality on pregnancy outcome in elderly women undergoing in vitro fertilization embryo transfer (IVF ET). J In Vitro Fert Embryo Transf 7:257-261, 1990.

Fisher SJ, Gimenez T, Henricks DM: Immunosuppressive activity associated with early pregnancy in the bovine. Biol Reprod 32:894, 1985.

Franchimont P, Hazee Hagelstein MT, Hazout A, Frydman R, Schatz B, Demerle F: Correlation between follicular fluid content and the results of in vitro fertilization and embryo transfer. I. Sex steroids. Fertil Steril 52:1006-1011, 1989.

Freitas S, Le Gal F, Dzik A, Volante M, Lelaidier C, Olivennes F, Fanchin R, de Zeigler D, Hazout A, Frydman R: Value of Cryopreservation of human embryos during the blastocyst stage. Contracept Fertil Sex 22:396-401, 1994.

Gott AL, Hardy K, Winston RML, Leese HJ: Non-invasive measurement of pyruvate and glucose uptake and lactate production by single human preimplantation embryos. Hum Reprod 5:104-108, 1990.

Grillo JM, Gamerre M, Lacroix O, Noizet A, Vitry G: Influence of the morphological aspect of embryos obtaine by in vitro fertilization on their implantation rate. J In Vitro Fert Embryo Transf 8:317-321, 1991.

Grudzinskas JG, Gordon YB, Jeffrey D, ChardT: Specific and sensitive determination of pregnancy-specific β1-glycoprotein by radioimmunoassay A new pregnancy test. Lancet 1:331-335, 1977.

Hardy K, Hooper MAK, Handyside AH, Rutherford AJ, Winston RML, Leese HJ: Non invasive measurement of glucose and pyruvate uptake by individual human oocytes and preimplantation embryos. Hum Reprod 4:188-191, 1989.

Hill GA, Freeman M, Bastias MC, Rogers BJ, Herbert CM 3d, Osteen KG, Wentz AC: The influence of oocyte maturity and embryo quality on pregnancy rate in a program for in vitro fertilization embryo transfer. Fertil Steril. 52:801-806, 1989.

Hillier SG, Wickings EJ, Afnan M, Margara RA, Harlow CR, Winston RML: Intrafollicular sex steroids and oocyte function. In: Gamete quality and fertility regulation, Rolland R *et al* (eds), Elsevier Science Publishers BV, 1985.

Hoult IJ, de Crespigny LC, O'Herlihy C, Spiers AL, Lopata A, Kellow G, Johnston I, Robinson HP: Ultrasound control of clomiphene/human chorionic gonadotrophin stimulated cycles for oocyte recovery and in vitro fertilisation. Fertil Steril 36:316-319, 1981

Human Fertilisation and Embryology Authority (HFEA), Code of Practise, 1991 onwards; HFEA, Paxton House, 30 Artillery Lane, London E1 7LS, UK.

Johnson MR, Riddle AF, Irvine R, Sharma V, Collins WP, Nicolaides KH, Grudzinskas JG: Corpus luteum failure in ectopic pregnancy. Hum Reprod 8:1491-1495, 1993a.

Johnson MR, Riddle AF, Grudzinskas JG, Sharma JG, Collins WP, Nicolaides K H: Reduced circulating placental protein concentrations during the first trimester are associated with preterm labour and low birth weight. Hum Reprod 8:1942-1947, 1993b.

Jones KP, Warnock SH, Urry RL, Edwin SS, Mitchell MD: Immunosuppressive activity and alpha interferon concentrations in human embryo culture media as an index of potential for successful implantation. Fertil Steril 57:637-640, 1992

Jongbloet PH and Zwets JHJ: Preovulatory overripeness of the egg in the human subject. Int J Obstet Gynaecol 14:111-116, 1976.

Kerin JF, Warnes GM, Quinn P, Jeffrey R, Godfrey B, Broom TJ, McEvoy M, Kirby C, Johnson M, Cox LW: The effect of clomid induced superovulation on human follicular and luteal function for extracorporeal fertilization and embryo transfer. Clin Reprod Fertil 2:129-142, 1983.

Kreiner D, Liu HC, Itskovitz J, Veeck L, Rosenwaks Z: Follicular fluid estradiol and progesterone are markers of preovulatory oocyte quality. Fertil Steril 48:991-994, 1987.

Kruger TF, Acosta AA, Simmons KF, Swanson RJ, Matta JF, Veeck LL, Morshedi M, Brugo S: New method of evaluating sperm morphology with predictive value for human in vitro fertilization. Urology 30:248-251, 1987.

Kruger TF, Menkveld R, Stander FSH, Lombard CJ, Van der Merwe JP, van Zyl JA, Smith K: Sperm morphologic features as a prognostic factor in in vitro fertilization. Fertil Steril 46:1118-1123, 1986.

Laufer N, Tarlatzis BC, DeCherney AH, Behrman HR: Steroid secretion by human oocyte corona cumulus complexes associated with conceptions

following in vitro fertilization. J In Vitro Fertil Embryo Transf 2:156-161, 1985.

Lopata A, Hay DL: The potential of early human embryos to form blastocysts, hatch from their zona and secrete hCG in culture. Hum Reprod 4(Supplement):87-94, 1989.

Lopata A: Factors influencing the growth of human preimplantation embryos in vitro. In:Human conception in vitro, Edwards RG, Purdy JM (eds), N York Acad Press:207-211, 1982

Ma S, Kalousek DK, Zouves C, Yuen BH, Gomel V, Moon YS: The chromosomal complementsof cleaved human embryos resulting from in vitro fertilization. J In Vitro Fertil Embryo Transf 7:16-21,1990.

Matthews JNS, Altman DG, Cambell MJ, Royston P: Analysis of serial measurements in medical research. Br Med J 300:230-235, 1990.

Mayumi T, Bitoh S, Anan S, Hama T, Fujimoto S, Yamamoto H: Suppressor T lymphocyte induction by a factor released from cultured blastocysts. J Immunol 134:404-409, 1985.

Menezo YJ, Guerin JF, Czyba JC: Improvement of human early embryo development in vitro by coculture on monolayers of Vero cells. Biol Reprod 42:301-306,1990.

Michaeli G, Moshe Fejgin, Ghetler Y, Nun IB, Beyth Y, Amiel A: Chromosomal analysis of unfertilized oocytes and morphologically

abnormal preimplantation embryos from and in vitro fertilization program. J In Vitro Fertil Embryo Transf 7:341-346, 1990.

Michie D: Methodologies from Machine Learning in Data Analysis and Software: The Computer Journal 34:559-565, 1991

Michie D: Personal models of rationality: J. Statistical Planning and Inference 25:381-399, 1988.

Michie D: Quality control of induced rule based programs: The Fifth generation. The GS Institute, London, 1984.

Mohr LR, Trounson AO, Leeton JF, Wood C: Evaluation of normal and abnormal human embryo development during procedures in vitro. In: Fertilization of the human egg in vitro, Beier HM, Lindner HR (eds), Springer-Verlag Berlin Heidelberg: 211-221, 1983.

Mortimer D: The male factor in infertility Part II: Sperm function testing. In: Current problems in Obstetrics, Gynaecology and Infertility, Volume III, No 8, Year Book Medical Publishers, Chicago:1985.

Murray MK, Segerson EC, Hansen PJ, Bazer FW, Roberts RM: Suppression of lymphocyte activation by a high-molecular weight glycoprotein released from preimplantation ovine and porcine conceptuses. Am J Reprod Immunol Microbiol 14:38,1987

Nayudu PL, Gook DA, Hepworth G, Lopata A, Johnston WI: Prediction of outcome in human in vitro fertilization based on follicular and stimulation response variables. Fertil Steril 51:117-125, 1989.

Nayudu PL, Gook DA, Lopata A, Sheather SJ, Lloyd Smith CW, Cadusch P, Johnston WI: Follicular characteristics associated with viable pregnancy after in vitro fertilization in humans. Gamete Res 18:37-55, 1987.

Nijs M, Geerts L, van Roosendaal E, Segal Bertin G, Vanderzwalmen P, Schoysman R: Prevention of multiple pregnancies in an in vitro fertilization program. Fertil Steril 59:1245-1250, 1993.

O'Neill C, Saunders DM: Assessment of embryo quality. Lancet ii:1035, 1984.

Oehninger S, Acosta AA, Morshedi M, Veeck L, Swanson RJ, Simmons K, Rosenwaks Z: Corrective measures and pregnancy outcome in in vitro fertilization in patients with severe sperm morphology abnormalities. Fertil Steril 50:283-287, 1988.

Pellestor F, Dufour MC, Arnal F, Humeau C: Direct assessment of the rate of chromosomal abnormalities in grade IV human embryos produced by in vitro fertilization procedure. Hum Reprod 9:293-302, 1994.

Plachot M, Mandelbaum J, Junca AM, Cohen J, Salat-Baroux J, Lage CD: Morphologic and cytologic study of human embryos obtained by in vitro fertilization. In: Future aspects in human in vitro fertilization, Kemeter P, Feichtinger W (eds), Springer-Verlag Berlin Heidelberg:267-275, 1987a.

Plachot M, Junca A, Mandelbaumm J, de Grouchy J, Salat-Baroux, Cohen J: Chromosome investigations in early life. II. human pre-implantation embryos. Hum Reprod 2:29-35, 1987b.

Plachot M, de Grouchy J, Junca AM, Mandelbaum J, Salat-Baroux J, Cohen J: Chromosomal analysis of human oocytes and embryos: does delayed fertilization increase chromosome imbalance? Hum Reprod 3:125-127, 1988a

Plachot M, Veiga A, Montagu J, de Grouch J, Calderon G, Lepretre S, Junca A, Santalo J, Carles E, Mandelbaum J, Barri P, Degoy J, Cohen J, Egozcue J, Sabatier JC, Salat-Baroux J: Are clinical and biological IVF parameters correlated with chromosomal disorders in early life: a multicentric study. Hum Reprod 3:627-635, 1988b.

Plachot M: Choosing the right embryo: the challenge of the nineties. J In Vitro Fert Embryo Transf 6:193-194,1989.

Pope VZ, Pope CE, Beck LR: SP-1 secretion by baboon embryos in vitro. Placenta 5:403-412, 1984.

Puissant F, Rysselberge MV, Barlow P, Deweze J, Leroy F: Embryo scoring as a prognostic tool in IVF treatment. Hum Reprod 2:705-708, 1987.

Punjabi U, Vereecken A, Delbeke L, Angle M, Gielis M, Gerris J, Johnston J, Buytaert PP: Embryo derived platelet activating factor, a marker of

embryo quality and viability following ovarian stimulation for in vitro fertilization. J In Vitro Fert Embryo Transf 7:321-326, 1990.

Quinlan, JR: C4.5 Programs for Machine Learning, Morgan Kaufmann. 1993.

Raman L, Pawashe AB: Plasma pregnancy specific β1-glycoprotein (SP1) in pregnancy. Indian J Med Res 84:223-226, 1986.

Saith RR and Sargent IL: Embryo selection for transfer in human IVF. Invited review, Assisted Reproduction Reviews. 5:3, 145-154, 1995.

Saith RR, Bersinger NA, Barlow DH and Sargent IL:. The role of Pregnancy-Specific Beta-1-Glycoprotein (SP1) in assessing human blastocyst quality in vitro. Hum Reprod (In press)

Saith RR, Srinivasan A, Sargent IS and Barlow DH: Embryo selection: Can artificial intelligence help? 10th Annual meeting of the European Society of Human Reproduction and Embryology, Belgaum, Poster abstract, 1994.

Saith RR, Sargent IL, Dokras A and Barlow DH: Immunosuppressive activity as a marker of human blastocyst quality in vitro. 9th Annual meeting of the European society of Human Reproduction and Embryology at Thessaloniki, Greece. Poster, 1993.

Scott RT Jr, Hofmann GE, Veeck LL, Jones HW Jr, Muasher SJ: Embryo quality and pregnancy rates in patients attempting pregnancy through in vitro fertilization. Fertil Steril. 55:426-428, 1991.

Scott RT, Hofmann GE, Muasher SJ, Acosta AA, Kreiner DK, Rosenwaks Z.: Correlation of follicular diameter with oocyte recovery and maturity at the time of transvaginal follicular aspiration. J In Vitro Fert Embryo Transf 6(2):73-75, 1989.

Segars JH, Niblack GD, Osteen KG, Rojers BJ, Wentz AC: The human blastocyst produces a soluble factor(s) that interferes with lymphocyte proliferation. Fertil Steril 52:381-387, 1989.

Sheth KV, Roca GL, Al Sediary ST, Parhar RS Hamilton CJ Al Abdul Jabbar F: Prediction of successful embryo implantation by measuring interleukin-1-alpha and immunosuppressive factor(s) in preimplantation embryo culture fluid. Fertil Steril 55:952-957, 1991.

Simonetti S, Veeck LL, Jones HW Jr: Correlation of follicular fluid volume with oocyte morphology from follicles stimulated by human menopausal gonadotropin. Fertil Steril 44:177-180, 1985.

Sinosich MJ, Grudzinskas JG, Saunders DM: Placental proteins in the Diagnosis and Evaluation of the "Elusive" Early Pregnancy. Obstet and Gynaaecol Survey 40:273-282, 1985.

Smith EM, Anthony F, Masson GM: Follicular fluid prostaglandin E_2:prostaglandin F2 alpha-ratio in relation to outcome of matched oocyte. Hum Reprod 6:519-523, 1991.

SPSS Advanced Statistics User's Guide, 1990.

SPSS for Windows, Advanced Statistics Release 6.0.

SPSS for Windows: Base system user's guide, Release 6.0. SPSS Inc. U.S.A. Descriptive statistics:167-180, 1993.

Staessen C, Camus M, Bollen N, Devroey P, Van Steirteghem AC: The relationship between embryo quality and the occurrence of multiple pregnancies. Fertil Steril 57:626-30, 1992.

Staessen C, Janssenswillen C, Van den Abbeel E, Devroey P, Van Steirteghem AC: Avoidance of triplet pregnancies by elective transfer of two good quality embryos. Hum Reprod 8:1650-1653, 1993.

Steer CV, Mills CL, Tan SL, Campbell S, Edwards RG: The cumulative embryo score: a predictive embryo scoring technique to select the optimal number of embryos to transfer in an in-vitro fertilization and embryo transfer programme. Hum Reprod 7:117-119, 1992.

Tarlatzis BC, Laufer N, DeCherney AH, Polan ML, Haseltine FP, Behrman HR: Adenosine 3',5' monophosphate levels in human follicular fluid: relationship to oocyte maturation and achievement of pregnancy after in vitro fertilization. J Clin Endocrinol Metab 60:1111-1115, 1985.

Testart J, Frydman R, Castanier M, Lassalle B, Belaisch J: Influence of ovarian stimulation on follicular fluid steroid levels and fertilization of the oocyte. In: Fertilization of the human egg in vitro, Beier HM, Lindner HR (eds), Springer-Verlag Berlin Heidelberg:73-82, 1983.

Testart J: Cleavage stage of human embryos two days after fertilization in vitro and their developmental ability after transfer into the uterus. Hum Reprod 1:29-31, 1986.

Tournaye H, Devroey P, Camus M, Staessen C, Bollen N, Smitz J, Van Steirteghem AC: Comparison of in vitro fertilization in male and tubal infertility a 3 year survey. Hum Reprod 7:218-222, 1992.

Trounson A: Factors controlling normal embryo development and implantation of human oocytes fertilized in vitro. In: Fertilization of the human egg in vitro, Beier HM, Lindner HR (eds), Springer-Verlag Berlin Heidelberg:235-250, 1983.

Trounson AO, Leeton JF, Wood C, Webb J, Wood J: Successful human pregnancies by in vitro fertilisation and embryo transfer in the controlled ovulation cycle. Science 212:681-682, 1981.

Turner K, Martin KL, Woodward BJ, Lenton EA, Leese HJ: Comparison of pyruvate uptake by embryos derived from conception and non-conception natural cycles. Hum Reprod 9:2362-2366, 1994.

Van Uem JF, Acosta AA, Swanson RJ, Mayer J, Ackerman S, Burkman LJ, Veeck L, McDowell JS, Bernardus RE, Jones HW Jr: Male factor evaluation in in vitro fertilization: Norfolk experience. Fertil Steril 44:375-383, 1985.

Van-Vlasselaer P, Vandeputte M: Immunosuppressive properties of murine trophoblast. Cell Immunol 83:422-432, 1984.

Vauthier-Brouzes D, Lefebvre G, Lesourd S, Gonzales J, Darbois Y: How many embryos should be transferred in in vitro fertilization? A prospective randomized study. FertilSteril 62:339-342, 1994.

Veeck LL: Oocyte assessment and biological performance. Ann NY Acad Sci 541:259-274, 1988.

Waterstone J, Parsons J, Bolton V: Elective transfer of two embryos. Lancet 337:975-976, 1991

Watson AJ: The cell biology of blastocyst development. Mol Reprod Dev 33:492-504, 1992.

Weiss SM, Kulikowski CA: Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning and expert systems. Morgan Kaufmann, San Mateo, 1991.

Winston NJ, Braude PR, Pickering SJ, George MA, Cant A, Currie J, Johnson MH: The incidence of abnormal morphology and nucleocytoplasmic ratios in 2-, 3- and 5-day human pre-embryos. Hum Reprod 6:17-24,1991.

Yovich JL: Embryo quality and pregnancy rates in in-vitro fertilisation. Lancet:283-284, 1985.

Zenzes MT, Belkien L, Bordt J, Kan I, Schneider HP, Nieschlag E: Cytologic investigation of human in vitro fertilization failures. Fertil Steril 43:883-891, 1985

Zhang WY, Teng H, Wang LY, Yan GL: Serum SP1, HPL levels and urinary estrogen/creatinine ratio in the detection of low birth weight. Chin Med J Engl 106:437-470, 1993.

Zubrick S: Validation of a weather forcasting expert system. Machine Intelligence 11. Eds: J. Hayes, D. Michie and J. Richard, pp. 391-422. OUP, Oxford, 1988.