THE USE OF NOVEL TUMOUR MARKERS AND STATISTICAL MODELS IN THE PREOPERATIVE DIAGNOSIS OF OVARIAN CANCER

Alexandra Claire Lawrence

MD

The University of Edinburgh

2007

Contents

Abstract		**9**
Declaration		**11**
Acknowledgements		**12**
Abbreviations		**13**
Chapter 1:	**Literature Review**	**15**
1.1	**Epidemiology and importance of ovarian cancer**	**16**
1.2	**The management of women with an adnexal mass**	**21**
1.3	**The diagnosis of ovarian cancer**	**25**
1.4	**Ultrasound**	**27**
1.4.1	The physics of gray scale ultrasound	27
1.4.2	The physics of Doppler	29
1.4.3	History of ultrasound	33
1.4.4	Morphology	34
1.4.5	IOTA classification of morphology	42
1.4.6	Subjective assessment	46
1.4.7	Doppler assessment	46
1.5	**Serum Tumour Markers**	**51**
1.5.1	CA125	51
1.5.2	CA19.9	54
1.5.3	CA72.4	58
1.5.4	CA15.3	60
1.5.5	Her-2/neu	62
1.6	**Statistical models**	**65**
1.6.1	Risk of malignancy indices	65
1.6.2	Logistic regression models	67
1.6.3	Artificial neural networks	70
1.6.4	Decision trees	75
1.6.5	Validation of prognostic models	77
1.6.6	Prospective validation of models	79
1.6.7	Small masses	83
1.7	**Hypotheses**	**86**
1.8	**Aims of the Thesis**	**87**
Chapter 2:	**General Methods**	**88**
2.1	**Setting: The Early Pregnancy & Gynaecology Assessment Unit (EPAGU), King's College Hospital, London**	**89**
2.2	**Subjects**	**91**
2.3	**Data collection**	**93**
2.4	**Ultrasound examination**	**94**

2.4.1	Morphology	95
2.4.2	Echogenicity	98
2.4.3	Septum	99
2.4.4	Papillary projection	100
2.4.5	Acoustic shadowing	103
2.4.6	Ascites	103
2.4.7	Colour doppler imaging	104
2.5	**Venous blood samples**	**107**
2.6	**Measurement of tumour markers**	**108**
2.7	**Histology**	**109**
2.8	**Statistical analysis**	**110**
2.9	**Prospective testing of published models**	**111**
2.10	**Decision tree analysis**	**114**
2.11	**Artificial neural network analysis**	**115**
2.12	**Ethics committee approval**	**117**
Chapter 3:	**The use of novel tumour markers to predict the presence of malignancy**	**118**
3.1	**Background**	**119**
3.2	**Methods**	**122**
3.2.1	Venous blood samples	122
3.2.2	Measurement of tumour markers	122
3.2.3	Artificial neural networks	122
3.3	**Results**	**124**
3.3.1	Histology	124
3.3.2	Tumor markers	125
3.3.3	Decision tree analysis	136
3.3.4	Artificial neural networks	139
3.4	**Discussion**	**144**
Chapter 4:	**A prospective evaluation of models for the diagnosis of ovarian cancer**	**148**
4.1	**Background**	**149**
4.2	**Population**	**150**
4.2.1	Inclusion Criteria	150
4.2.2	Exclusion Criteria	150
4.3	**Statistical analysis**	**151**
4.4	**Results**	**152**
4.4.1	Demographic variables	154
4.4.2	Ultrasound variables	158
4.4.3	Doppler variables	161
4.4.4	Diagnosis of borderline tumours	163
4.4.5	Subjective assessment of mass	164

4.4.6	Prospective testing of models	166
4.4.7	Artificial neural networks	169
4.5	**Discussion**	**175**
4.5.1	Univariate analysis	175
4.5.2	Subjective assessment	176
4.5.3	Prospective evaluation	177
4.5.4	Artificial neural networks	180
Chapter 5:	**The application of models to incidentally found small masses**	**182**
5.1	**Background**	**183**
5.2	**Population**	**185**
5.3	**Results**	**186**
5.3.1	Histology of small masses	186
5.3.2	Presenting symptoms	187
5.3.3	Univariate analysis	187
5.3.4	Comparison of large and small masses	188
5.3.5	Subjective assessment	190
5.3.6	Prospective performance of statistical models	191
5.4	**Discussion**	**193**
Chapter 6:	**General Discussion**	**196**
6.1	**Introduction**	**197**
6.2	**Synopsis of results**	**198**
6.3	**Discussion: the wider context**	**203**
6.4	**Suggestions for future studies**	**205**
Bibliography		**207**
Appendices		**222**
I	International Federation of Gynaecology and Obstetrics (FIGO) staging for primary carcinoma of the ovary	222
II	Information sheet on the value of biochemical markers and ultrasound in the mamangement of ovarian cysts.	223
III	GP letter	224
IV	Consent form: Study of the value of biochemical markers and ultrasound in the management of ovarian cysts.	225
V	Measurement of tumour markers	226
VI	Conference proceedings	231

List of Tables

Chapter 1

1.1	Epidemiology of the most common cancers in the United Kingdom in 2002	17
1.2	Survival of women with epithelial cancer by stage and the percentage of women diagnosed in each stage	20
1.3	Sensitivity, specificity and likelihood ratio for malignancy given a positive or negative result for different levels of the risk of malignancy index	67
1.4	Comparison of the performance of different models	70
1.5	Performance of Jacobs' and Tingulstad's RMIs at different cut off levels	80
1.6	Prospective performance of Jacobs' RMI	81
1.7	Prospective performance of logistic regression models	82
1.8	Prevalence of adnexal masses and primary ovarian cancer in asymptomatic women screened using ultrasound	84
1.9	Scoring system for small masses	85
1.10	Comparison of the performance of Ferazzi's scoring system to other morphological scores on small masses	85

Chapter 2

2.1	Reproducibility of blood flow indices with CDI and CDE	105
2.2	Prospectively tested models: a summary of their equations, variables and year of publication	112

Chapter 3

3.1	Histology of masses	124
3.2	Values of CA 125, CA 72-4, CA 15-3, CA 19-9 and Her-2/neu with different histology	127
3.3	Area under the ROC curve for tumour markers and age	134
3.4	Values of CA 125, CA 72-4, CA 15-3, CA 19-9 and Her-2/neu with different benign histologies	135
3.5	Values of CA 125, CA 72-4, CA 15-3, CA 19-9 and Her-2/neu with different invasive histologies	136
3.6	Comparison of the performance of TNN1 and TNN2	143

Chapter 4

4.1	Histological classification of the adnexal masses	153
4.2	Comparison of the demographic variables to histology	156
4.3	Comparison of presenting symptom to histology	157
4.4	Severity of pain	158

4.5	Distribution of size and volume in benign and malignant masses	159
4.6	Locularity of adnexal masses	160
4.7	Echogenicity of cyst fluid	160
4.8	Presence of binary ultrasound variables in benign and malignant masses	161
4.9	Doppler variables in benign, borderline and invasive masses	162
4.10	Distribution of colour score	162
4.11	Significant variables in the discrimination between benign, borderline and invasive masses	163
4.12	Comparison of subjective impression to histology	165
4.13	Univariate analysis of all binary end points	166
4.14	Comparison of the prospective performance of different logistic regression models	168
4.15	Significant variables in each training set	170
4.16	Comparison of the performance of all neural networks	174

Chapter 5

5.1	Histology and stage of small adnexal masses	186
5.2	Presenting symptoms of small masses	187
5.3	Demographic and morphological variables by histology of mass	188
5.4	Demographic and morphological variables in large and small masses	189
5.5	Presenting symptoms of large and small masses	189
5.6	Locularity of larger and small masses	190
5.7	Accuracy of subjective assessment in the specific diagnosis of masses	190
5.8	Performance of models on small masses	192

Chapter 6

6.1	An overview of the performance of all models investigated in the thesis	201

Appendix V

A.1	Minimum sensitivities of cancer antigen assay kits	229
A.2	The intra and inter-assay coefficients of variation for cancer antigen assays	229

List of Figures

Chapter 1

1.1	Age specific incidence of all types of ovarian cancer in South Thames, 1995	19
1.2	Acoustic shadowing caused by a soft tissue/tooth interface in a dermoid cyst	29
1.3	Doppler waveform characteristics	32
1.4	A cystic corpus luteum within the ovary	35
1.5	A fimbrial cyst lying lateral to the ovary	36
1.6	A mucinous cystadenoma	37
1.7	A small solid dermoid placed laterally in the ovary surrounded by healthy ovarian tissue	38
1.8	A serous cystadenofibroma	40
1.9	A diastolic notch in the Doppler waveform	48
1.10	The sythesis pathway of the Lewis a antigen, the Lewis b antigen and CA 19-9	57
1.11	A simple neural network	73
1.12	A decision tree for the classification of irises	77

Chapter 2

2.1	Local Health Authorities in London and South East London	90
2.2	Examples of unilocular cysts	95
2.3	Examples of unilocular solid cysts	96
2.4	Examples of multilocular cysts	96
2.5	Examples of multilocular solid cysts	97
2.6	Examples of solid masses	98
2.7	Measurement of septum thickness	99
2.8	A serous cystadenoma consisting of a unilocular anechoic cyst with a single papillary projection	101
2.9	Measurement of a papillary projection	102
2.10	Measurement of the height of a papillary projection arising from a solid area	103
2.11	Measurement of free fluid in the pouch of Douglas	104

Chapter 3

3.1	Distribution of CA 125 with benign, borderline and invasive masses	128
3.2	Distribution of CA 72-4 with benign, borderline and invasive masses	129
3.3	Distribution of CA 15-3 with benign, borderline and invasive masses	130

3.4	Distribution of CA 19-9 with benign, borderline and invasive masses	131
3.5	Distribution of Her-2/neu with benign, borderline and invasive masses	132
3.6	Receiver Operator Characteristic curve for tumour markers and age against benign and malignant histology	133
3.7	Decision tree 1 for the discrimination between benign and malignant masses	138
3.8	Decision tree 2 for the discrimination between benign, borderline and invasive masses	139
3.9	Receiver Operator Characteristic curves for the two neural networks	142

Chapter 4

4.1	Subjective classification of the masses into specific histological types	165
4.2	Receiver Operator Characteristic curves for the ten morphological networks	171

Appendix V

A.1	The colorimetric reaction of *p*-nitrophenyl phosphate	227

Abstract

Malignant ovarian tumours are diagnosed at an advanced stage in 75% of cases and they have the highest mortality figures of all gynaecological cancers. As the treatment of benign and malignant adnexal masses is significantly different it is important to be able to reliably distinguish between them preoperatively. Thus, women with malignancies could be referred to cancer centres, whilst those with benign conditions could be offered more conservative management.

The aims of this thesis are (1) To investigate the use of new tumour markers in the preoperative diagnosis of ovarian cancer. (2) To validate previously published models and compare their performance to subjective assessment and to the models developed in this thesis. (3) To investigate the differences between small asymptomatic masses and large masses and to investigate the accuracy of published models on the diagnosis of malignancy in small masses.

CA 125, CA 15-3 and CA 72-4 were significantly raised in the presence of ovarian cancer. CA 72-4 was higher in mucinous cancers and CA 125 and CA 15-3 were higher in serous and endometrioid cancers. Her-2/neu and CA 19-9 were not significantly different in benign or malignant disease. Logistic regression analysis showed age, CA125 and CA 15-3 to be the most valuable discriminators. A neural network was designed and trained which gave a sensitivity of 100% and a specificity of 90.9% on the test set. None of the six published models tested prospectively performed as well as in their original publication. The IOTA logistic regression model performed best and gave a sensitivity of 81.8% and a specificity of 72.3%. Subjective assessment of the

mass gave a sensitivity of 72.7% with a specificity of 81.8%. Small masses were more commonly unilocular and large masses multilocular. Ascites, papillary proliferations, detectable flow and the smoothness of the internal wall discriminated well between benign and malignant small cysts. Age, menopausal status and CA125 were not discriminatory. None of the published models were as accurate as subjective assessment at diagnosing malignancy. These data suggest that statistical models may be of less value than tumour markers and subjective assessment in the diagnosis of ovarian malignancy.

This work improves our ability to predict malignancy in a pelvic mass. As a result of this work, further research might aim to combine the use of tumour markers and subjective assessment to improve the preoperative diagnosis of malignancy. It may thus be possible to provide care in a cancer centre for those women that need it and to allow conservative management or minimally invasive surgery for women with benign disease.

Declaration

Except where due acknowledgement is made by reference, the studies undertaken in this work were the unaided work of the author. No part of this work has been previously accepted for, or is currently being submitted in candidature for another degree.

Dr Alexandra Claire Lawrence

Acknowledgements

I would like to thank my supervisor, Davor Jurkovic for his contagious enthusiasm, support and encouragment. I am also grateful for him for teaching me how to scan and how to be a better clinician. He funded my two years as a research fellow and I thank him for his generosity.

My colleagues in the Early Pregnancy and Gynaecology Assessment Unit have given me support and help in recruiting patients: Janine Elson, Rehan Salim, Naaila Aslam and Brigitte Woelfer.

Dr Roy Sherwood and Ms Jenny Jones (DSL Laboratory, King's College Hospital) taught me how to run the tumour marker assays. My thanks to Dr Derek Cooper for statistical advice.

Finally, I would like to thank my family and friends for their unfailing support and encouragement.

Abbreviations

ALP	alkaline phophatase
AFP	alpha fetoprotein
ANN	artificial neural network
AUC	area under the curve
CA 125	cancer antigen 125
CA 15-3	cancer antigen 15-3
CA 19-9	cancer antigen 19-9
CA 72-4	cancer antigen 72-4
CART	classification and regression tree
CDE	colour Doppler energy
CDI	colour Doppler imaging
CEA	carcinoembryonic antigen
CT	computed tomogram
D	end diastolic velocity
ECD	extracellular ligand-binding domain
EPAGU	early pregnancy and gynaecology assessment unit
Fab	fragment antigen binding
IOTA	international ovarian tumour analysis group

LRM	logistic regression model
Mab	monoclonal antibody
MRI	magnetic resonance imaging
NN	neural network
NPV	negative predictive value
TAMXV	time averaged maximum velocity
PI	pulsatility index
PPV	positive predictive value
PRF	pulse repetition frequency
PSV	peak systolic velocity
RI	resistance index
RMI	risk of malignancy index
ROC	receiver operator characteristic
S	peak systolic velocity

Chapter 1

Literature Review

1.1 Epidemiology and importance of ovarian cancer

Ovarian cancer is the biggest challenge facing gynaecological oncologists at the present time. Epithelial ovarian malignancy is the most common ovarian cancer and, due to its non-specific symptoms, the diagnosis is delayed until the disease is advanced in over 70% of cases. Ovarian cancer affects around one percent of women in the United Kingdom and has the highest fatality to case ratio of all gynaecological malignancies (World Health Organisation, 2002). It is the fourth most common cause of death from malignancy in women after lung, breast and colon cancers (Quinn et al 2001) (Table 1.1). Once contracted, there is a 62% risk of dying from ovarian cancer as compared to a 30% risk from breast cancer and 20% from endometrial cancer in Western Europe. The mortality of ovarian cancer in the UK is 40% higher than other developed countries within Europe and is the third highest in the Eur-A countries* (Highlights on health, United Kingdom 2004, WHO). It accounts for 6% of all cancer-related deaths in women in the UK, more than all the other gynaecological cancers combined. The surgical treatment of ovarian cancer is demanding and requires both physical and psychological resilience from the patient.

*The group comprises Andorra, Austria, Belgium, Croatia, Cyprus, the Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Monaco, the Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Sweden, Switzerland and the United Kingdom.

Table 1.1: Epidemiology of the most common cancers in the United Kingdom in 2002 (International Agency for Research on Cancer, Lyon, France)

	Cases	Crude Rate/100000	Deaths	Crude Rate/100000
Breast	40928	135.5	13303	44.0
Colon and rectum	16562	54.8	8278	27.4
Lung	15424	51.0	13390	44.3
Ovary	6707	22.2	4590	15.2
Corpus uteri	5956	19.7	1183	3.9
Non-Hodgkin lymphoma	4409	14.6	2260	7.5
Melanoma of skin	3959	13.1	807	2.7
Pancreas	3720	12.3	3596	11.9
Stomach	3675	12.2	2719	9.0
Bladder	3445	11.4	1683	5.6
Cervix uteri	3181	10.5	1529	5.1

The incidence of ovarian cancer is significantly lower in sub-Saharan Africa than in Europe and the USA. The age-standardised rate is 6.6 per 100,000 in Uganda compared with 12.4 in the UK (Parkin, 1997). The highest rates are seen in Caucasian women in North America and Western Europe. Women living in China and Japan have lower rates, a difference preserved with immigration to the USA (Weiss, 1978).

The incidence of ovarian cancer shows two peaks in age, a small peak in the late teenage years and a second at the age of 80 (Quinn, 2001). This reflects

the two main types of ovarian cancer: non-epithelial and epithelial. Approximately 90% of ovarian tumours are epithelial in origin. This occurs more commonly in older women with a peak incidence at 56 to 60 years. The age-specific incidence continues to increase until the age of 80 when it slowly declines (Figure 1.1). The peak incidence of non-epithelial tumours is between the ages of 15 and 25 years. Non-epithelial tumours comprise tumours originating from germ cells, sex cord-stromal cells, ovarian metastases and rare cancers such as sarcoma and small cell carcinoma. Germ cell tumours make up two thirds of ovarian cancer presenting in the first two decades of life but are rare over the age of 40.

The survival of women is markedly different with epithelial and non-epithelial cancers. Epithelial tumours are relatively slow growing and are usually diagnosed at an advanced stage. The usual presenting symptoms such as abdominal pain or swelling, change in bowel habit, weight loss and vaginal bleeding are non-specific and are commonly experienced by a large number of the normal population (Flamm, 1988). Survival of women with epithelial cancer is significantly altered by both age and the stage and grade of disease. For all stages, women over the age of 50 have a 5-year survival of 15% whereas women below the age of 50 have a 40% 5-year survival. The data for survival and percentage of women diagnosed at each stage is shown in Table 1.2.

Figure 1.1: Age specific incidence of all types of ovarian cancer in South Thames, 1995

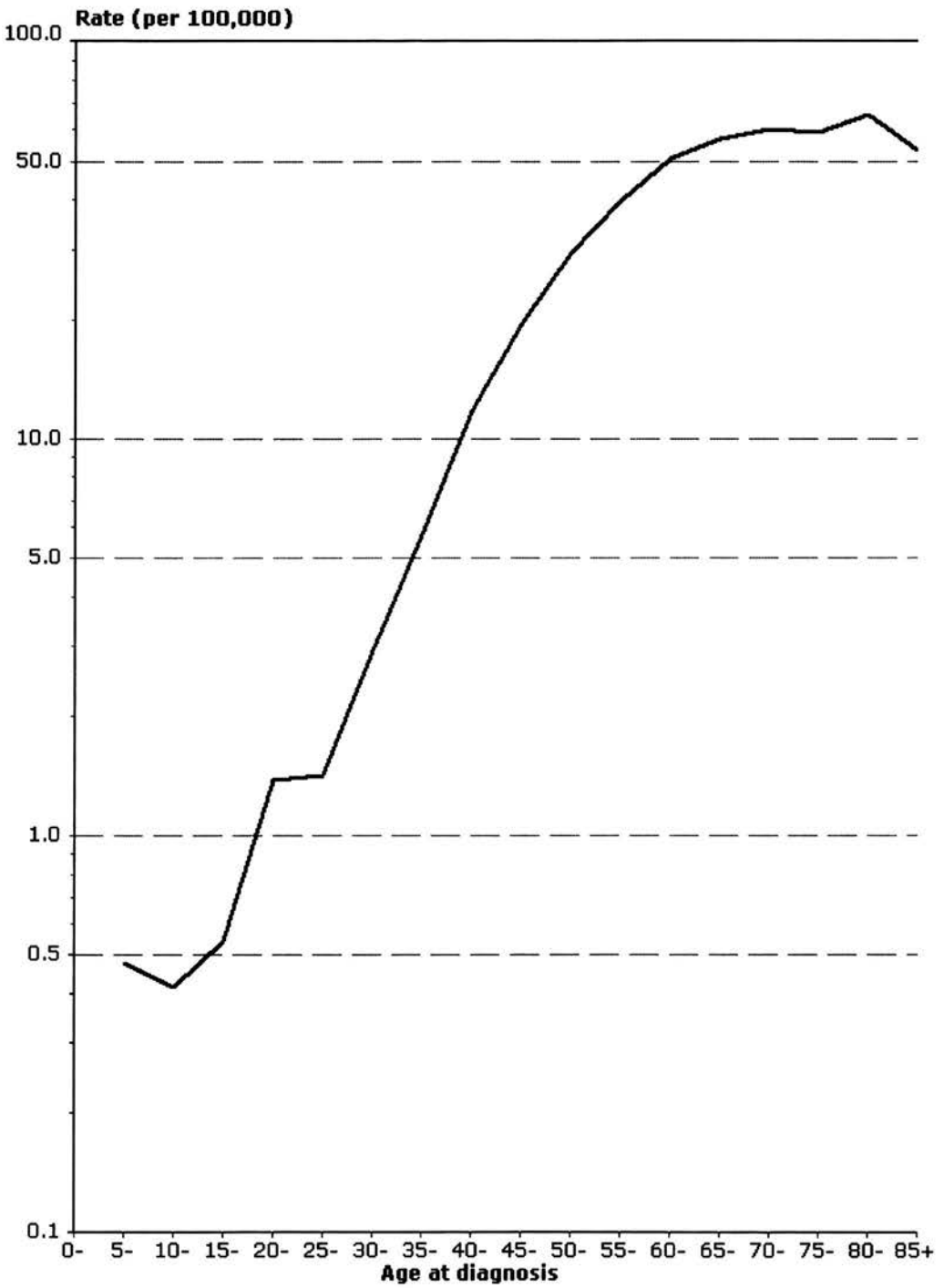


Table 1.2: Survival of women with epithelial cancer by stage and the percentage of women diagnosed in each stage

	Stage I Confined to ovaries	**Stage II** Confined to pelvis	**Stage III** Positive nodes or peritoneal implants	**Stage IV** Distant metastases
Stage at diagnosis (%)	21	7	63	9
5-year survival (%)	82	60-74	23-41	11

Non-epithelial tumours and specifically germ cell tumours grow rapidly and usually present early with symptoms of pain secondary to capsule distension, haemorrhage, necrosis or torsion. Germ cell tumours mainly present in Stage I; 65% of dysgerminomas are Stage I at diagnosis and 90% of these are unilateral (Bjorkholm et al, 1990). Survival of women with non-epithelial tumours is significantly better then those with epithelial tumours. This is due to the early age and stage of presentation and the high sensitivity of germ cell tumours to chemotherapy (Williams et al, 1994).

1.2 The management of women with an adnexal mass

The appropriate management of a woman presenting with a mass is dependent on the ability to distinguish between a physiological, benign and malignant tumour. A physiological tumour can be simply observed, in the absence of significant symptoms, until it resolves spontaneously (Soutter, 2003). A benign tumour can be treated expectantly in the case of a small, asymptomatic cyst. This is supported by work by Valentin in 2002. 134 postmenopausal women with asymptomatic, adnexal cysts of benign appearance were followed up with ultrasound for up to 8 years. 9% of women underwent surgery due to an increase in size or morphology and all were found to be benign. The remaining 91% had either static or regressing cysts for the duration of follow-up (Valentin, 2002). If surgery is required, laparoscopic surgery allows a swift recovery to normal activities.

A malignant tumour should be staged and optimally debulked by a gynaecological oncologist. A retrospective study of all women presenting with ovarian cancer in the West of Scotland showed that survival was significantly improved if a gynaecologist rather than a general surgeon performed the surgery (Junor et al, 1987). Life expectancy was also prolonged if the tumour was optimally debulked and if postoperative care was carried out by a joint clinic. A prospective observational study demonstrated a survival advantage for women managed by a subspecialist in gynaecological oncology. Subspecialists were found to perform adequate and appropriate surgery and to provide suitable adjuvant therapy more often than a generalist (Kehoe et al, 1994).

In view of these findings and those for other cancers, the Calman-Hine Report, *A Policy Framework for Commissioning Cancer Services,* was published in 1995. This proposed a plan for the management of cancer involving three service levels: primary care; Cancer Units and Cancer Centres to improve the outcome of people suspected of having cancer. A Cancer Unit is a team involving a lead gynaecologist, a lead pathologist, a radiologist and a nurse with a special interest in gynaecological cancer. It may be located in either one or more collaborating district general hospitals serving a population of 200,000 people. A Cancer Centre is staffed by a multiprofessional gynaecological oncology team. This would include two gynaecological oncologists, a clinical oncologist, a medical oncologist, a radiologist, a histopathologist, a cytopathologist and a clinical nurse specialist for a population of 1 million.

These recommendations were reinforced by the publication of *Improving Outcomes in Gynaecological Cancers* in 1999. This manual set out steps for the appropriate management of women suspected of having ovarian cancer. A woman presenting with symptoms or signs suggestive of early ovarian cancer would be referred to her local Cancer Unit. There she would undergo abdominal and pelvic examination, transvaginal ultrasound and CA 125 measurement. If these findings, along with her age, were suggestive of a malignant mass then she should be referred without delay to a multidisciplinary gynaecological oncology team at a Cancer Centre.

A gynaecological oncologist should carry out surgery through a midline incision and any free fluid should be aspirated and submitted for cytological investigation. In the absence of free fluid, peritoneal washings should be taken. The ovarian tumour should be removed intact if possible. An

international retrospective study of 1545 women with Stage I ovarian cancer identified intraoperative cyst rupture as an independent poor prognostic indicator for survival (Vergote et al, 2001). The contralateral ovary and uterus should be removed and a systematic inspection of all visceral surfaces and viscera should be conducted. Any suspicious areas should be biopsied and an infracolic omentectomy performed. For full FIGO staging, a diaphragmatic smear and exploration of the pelvic and paraaortic lymph nodes is also required.

The management of a woman suspected of having an ovarian cancer differs markedly from that of a tumour thought to be benign. The former would require a staging laparotomy at a hospital geographically distant from her home with the attendant risks of major abdominal surgery e.g. infection, visceral injury, the need for blood transfusion, deep venous thrombosis and perioperative death. The latter would have a choice of expectant or minimally invasive day surgery. There is also a considerable difference in the psychological impact of the diagnosis. The woman with suspected benign disease would have a choice of management in many cases and treatment would take place at her local hospital. She can choose to preserve her uterus and contralateral ovary and thus her fertility. The woman with suspected cancer would have the psychological burden of this diagnosis. She would have little choice over the type of surgery required and fertility-sparing surgery would be less feasible. The majority of gynaecological cancer patients have chronic sexual problems following surgery. 31% of women describe anxiety and 41% are depressed after gynaecological cancer surgery (Corney et al, 1992). In addition, treatment may take place further from home, friends and relatives increasing the sense of isolation and displacement.

Treatment is not required for functional ovarian cysts save in exceptional cases of severe pain. Indeed surgical intervention in many cases can increase morbidity and impair fertility.

Incidentally diagnosed pelvic masses are increasing in frequency due to the increased availability of diagnostic imaging in primary and secondary care. A large number of these asymptomatic women are thought to have an ovarian malignancy and undergo the significant surgery with its attendant physical and psychological risks (Timmerman et al, 1999a). Accurate preoperative discrimination between benign and malignant disease would particularly benefit this healthy population.

1.3 The diagnosis of ovarian cancer

In most women, the diagnosis of ovarian cancer is made on histological specimens taken at exploratory laparotomy.

Symptoms suggestive of ovarian cancer are often vague and non-specific, particularly those of epithelial ovarian cancer. Irregular vaginal bleeding may occur in early stage disease. If a pelvic mass is compressing the bladder or rectum, local symptoms of urinary frequency or constipation may be noted. Ascites and omental deposits can lead to abdominal distension, bloating and abdominal pain. Increased intra-abdominal pressure may cause nausea, anorexia or early satiety. Breathlessness may occur in late stage disease due to a malignant pleural effusion or diaphragmatic splinting. Non-epithelial tumours are faster growing and are more likely to present with pain due to haemorrhage or capsule rupture (Berek, 2005).

The most important sign of ovarian malignancy is a pelvic mass. An irregular, fixed pelvic mass is highly suggestive of cancer, particularly if associated with shifting dullness and an upper abdominal mass representing ascites and an omental cake. As women often report only abdominal symptoms, examination may not include a vaginal examination and a pelvic mass may be missed. The groin and neck examination may reveal enlarged lymph nodes (Soutter, 2003).

Investigations include routine haematological and biochemical measurements. Elevated liver enzymes or abnormal clotting times are suggestive of extensive liver metastases. Specific tumour markers are

discussed below. The primary radiological investigation is a pelvic ultrasound. Other preoperative imaging includes a chest radiograph to exclude lung metastases. Magnetic resonance imaging can identify the planes surrounding a tumour and the presence of tumour invasion into neighbouring structures. Computed tomography (CT) is useful in assessment of extraovarian disease, in particular the presence of an omental cake, diaphragmatic disease and implants in bowel mesentery. A recent study investigated the use of CT in the preoperative prediction of suboptimal cytoreduction in advanced ovarian cancer (Axtell, 2007). Diaphragmatic disease and large bowel mesentery implants were found to be most predictive of suboptimal cytoreduction. However, the high accuracy of prediction was not confirmed on external validation at two other US centres.

An extraovarian primary should be carefully excluded prior to planning surgery. A barium enema or colonoscopy will identify the presence of a colonic primary in patients with rectal bleeding or a change in bowel habit. Women with a breast lump should undergo mammography to exclude a primary breast cancer with ovarian metastasis.

1.4 Ultrasound

1.4.1 The physics of gray scale ultrasound

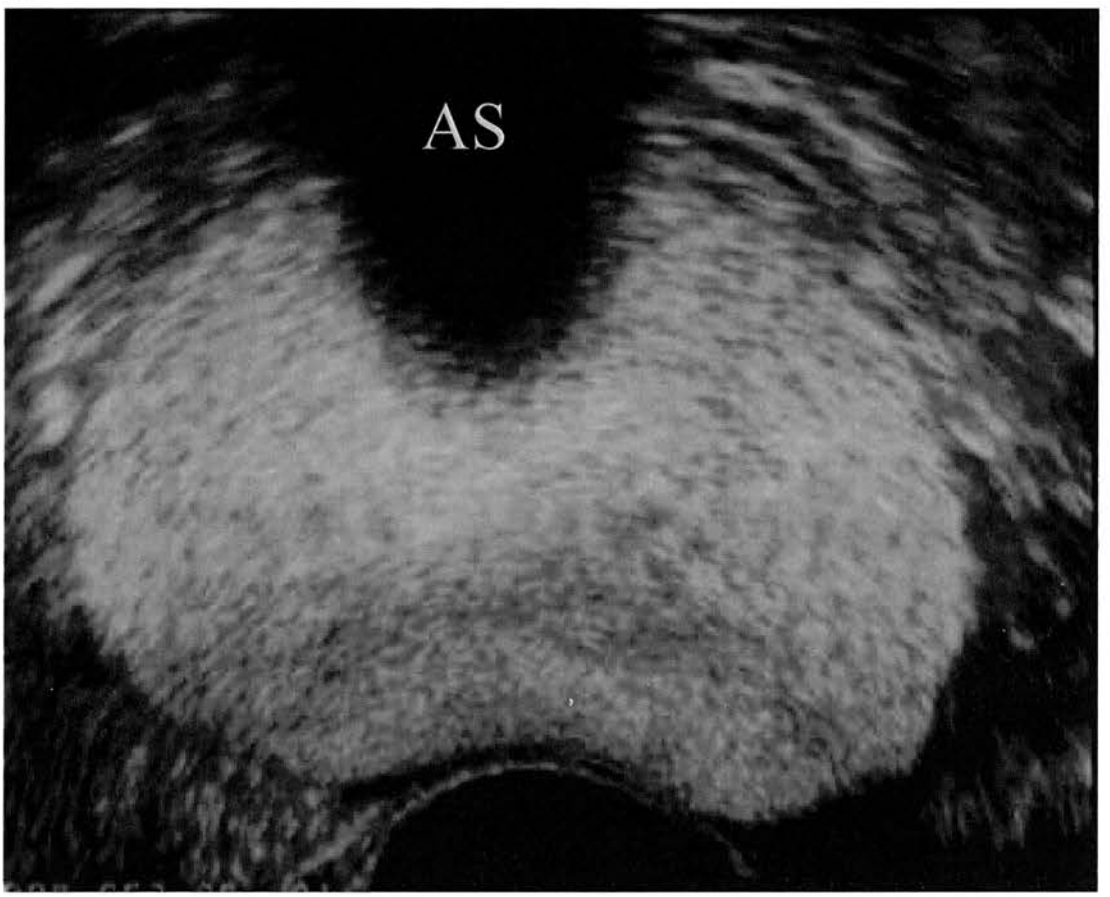
Ultrasound is reliant on the pulse echo principle. A pulse of sound waves emitted by a transmitter can be reflected by a tissue interface and these can be detected by a receiver. The elapsed time between emission and reception can be measured and the distance from the tissue interface can be calculated by the equation:

Distance = Speed x time

The speed of sound waves travelling in soft tissue is a known constant, 1540 metres/second, and is similar to that of fluid and blood. Modern ultrasound machines have transducers that combine a transmitter and receiver. This is constructed from a piezoelectric crystal that changes thickness when a voltage is applied across it. When the voltage is pulsed, the crystal resonates and produces high frequency sound depending on the thickness of the crystal. The crystal also acts as the receiver by producing a small electrical signal when hit by ultrasound waves. The crystal is used to produce bursts of sound every thousandth of a second, lasting a millionth of a second. The received echoes are quickly converted into an image allowing motion to be followed. This is known as real-time imaging. The image produced is that of a slice through a tissue. In order to examine an organ, the transducer is angled and moved to view a number of slices. Thus, there are no fixed projections: both the acquisition and interpretation of the images is dependent on of the skill of the operator.

Ultrasound frequencies for diagnostic use vary between 2.5 and 13 megahertz (MHz). Higher frequencies allow more definition of a structure but their tissue penetration is limited. The speed of sound in bone and air is markedly different from that of tissue. An interface between different tissues causes reflection of a portion of the sound waves. The greater the difference in speed of sound in the two tissues, the greater the proportion reflected. A tissue/fluid interface causes a small proportion of the sound to be reflected, the majority continuing through the tissues. This enables an image to be obtained of the interface and the tissue beyond. At a tissue/air interface there is a strong reflection: only one percent of the sound waves continue to penetrate the air. This gives a good view of the interface but minimal information about what lies beyond. This accounts for the inability of ultrasound to examine a bowel lumen and the strong reflection seen in a dermoid cyst containing a tooth or bone. This is known as acoustic shadowing (Figure 1.2).

Figure 1.2: Acoustic shadowing (AS) caused by a soft tissue/tooth interface in a dermoid cyst



1.4.2 The physics of the Doppler effect

The Doppler effect allows assessment of the vascularity of a structure. When an observer is moving relative to a wave source, the frequency measured is different to the emitted frequency. If the source and observer are moving towards each other, the observed frequency is higher than that emitted; if they are moving apart, it is lower. This was first described by Christian Andreas Doppler (1803 – 1853) in Austria (Doppler, 1843). This principle allows the detection of ultrasound reflected by red blood cells flowing in

tissues. The frequency shift can be measured by the transducer and blood flow velocity calculated by the formula:

$$\textbf{Frequency shift} = \frac{2 \times Fi \times v \times \cos\theta}{C}$$

where C is the speed of sound in tissue and Fi the incident frequency of the beam. Both C and Fi are constants. If the Doppler angle, θ, is also kept constant, the flow velocity v can be calculated from the frequency shift. The ultrasound transducer uses short pulses of sound, between which it remains silent to allow echo reception. The frequency of the pulses is described by the pulse repetition frequency (PRF).

Three types of Doppler assessment are available:

1.4.2.1 Colour flow Doppler

Colour flow Doppler allows an overview of blood flow within a region. It is represented as superimposed colour on the gray scale image. Blood moving towards the probe is shown in red, blood moving away in blue. Colour Doppler is affected by the angle of insonation (θ) and is limited by the PRF. The maximum frequency shift that can be resolved is half the PRF, known as the Nyquist limit. Frequencies higher than this will "wrap around" the frequency scale and appear as frequency shifts in the opposite direction, a phenomenon known as aliasing. Increasing the PRF in order to measure high frequency shifts as seen in arterial flow will reduce the effective penetration of the transducer.

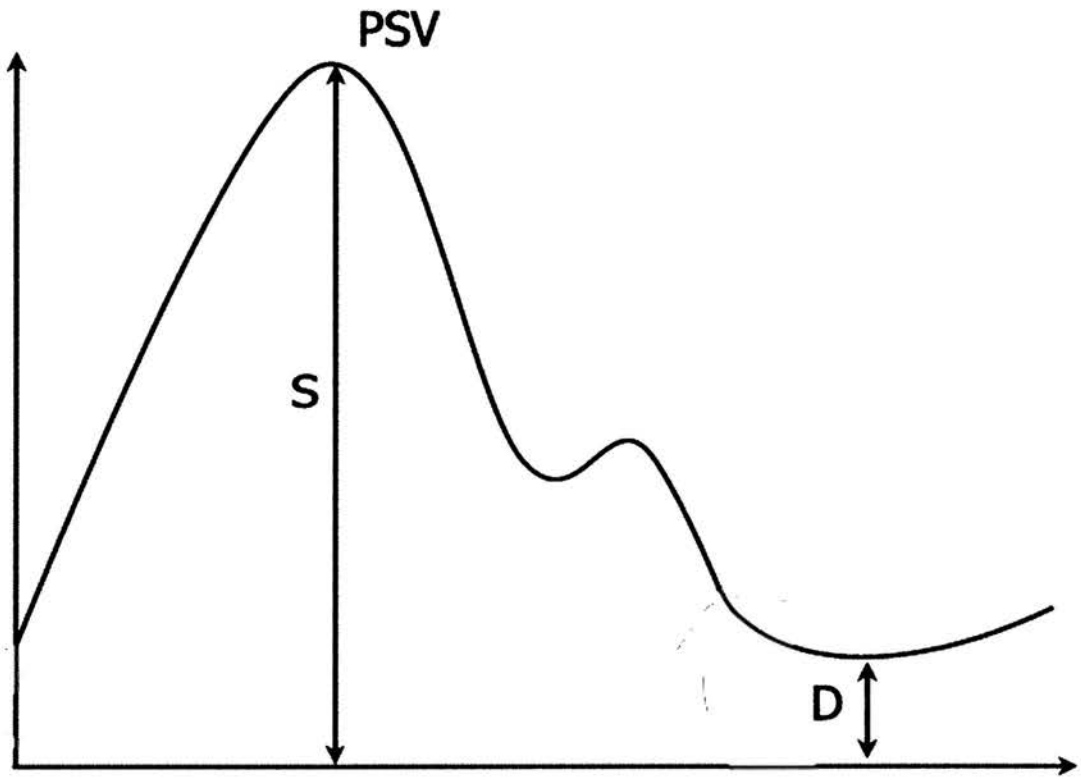
1.4.2.2 Power wave Doppler

Spectral or power wave Doppler enables the frequency shift for a small area to be observed. The area is selected by a gate positioned at will on the colour flow image. The Doppler signal obtained is plotted against time to give a flow velocity waveform. Each vessel has a specific waveform that is dependent on the characteristics of the vascular bed beyond; the recoil of the vessel wall and the resistance of the vessels upstream. Arteries supplying high resistance structures show absent or reversed flow at the end of diastole. Those supplying low resistance vascular beds have persistent flow throughout diastole. The waveform can be analysed to enable comparison between vessels in the same subject and between different subjects. The envelope of the waveform is outlined allowing the ultrasound machine to calculate waveform indices. Commonly used indices are peak systolic velocity (PSV), time-averaged maximum velocity (TAMXV), pulsatility index (PI) and resistive index (RI). The peak systolic velocity is the maximum observed velocity of the waveform (see Figure 1.3). The values S (also known as peak systolic velocity) and D (end diastolic velocity) are used in the calculation of the waveform indices using the equations:

$$\mathbf{RI} = \frac{\mathbf{S - D}}{\mathbf{S}} \qquad \mathbf{PI} = \frac{\mathbf{S - D}}{\mathbf{mean}}$$

The RI takes values between zero and one, whereas the PI runs from zero to infinity. The pulsatility index is able to reflect reversed end diastolic flow because the mean flow velocity is the denominator. The Doppler variables PI and RI were first described in the mid-1970's (Gosling and King, 1975, Pourcelot 1974).

Figure 1.3: Doppler waveform characteristics. PSV and S denote peak systolic velocity, D is end diastolic velocity.



1.4.2.3 Power Doppler

The third type of Doppler is power Doppler. This was introduced to overcome some of the limitations of conventional Doppler such as dependence on the angle of insonation (θ). Power Doppler measures the energy of moving red cells, which is dependent on the amplitude of the Doppler signal, the density of the red blood cells and the attenuation of the intervening tissue. This is displayed as colour on the gray scale image. The advantages of power Doppler are that it is less affected by artefacts; is independent of θ and is more sensitive to low flow than conventional Doppler. It is good at delineating the course and calibre of vessels. Its

limitations are that it does not show direction of flow or information on blood flow velocity.

1.4.3 History of ultrasound

The piezoelectric effect was discovered by the Curie brothers in France in 1880. They observed that an electric potential would be produced when mechanical pressure was exerted on a quartz crystal such as Rochelle salt (sodium potassium tartrate tetrahydrate). The reciprocal behaviour of achieving a mechanical stress in response to a voltage difference was mathematically deduced from thermodynamic principles by the physicist Gabriel Lippman in 1881.

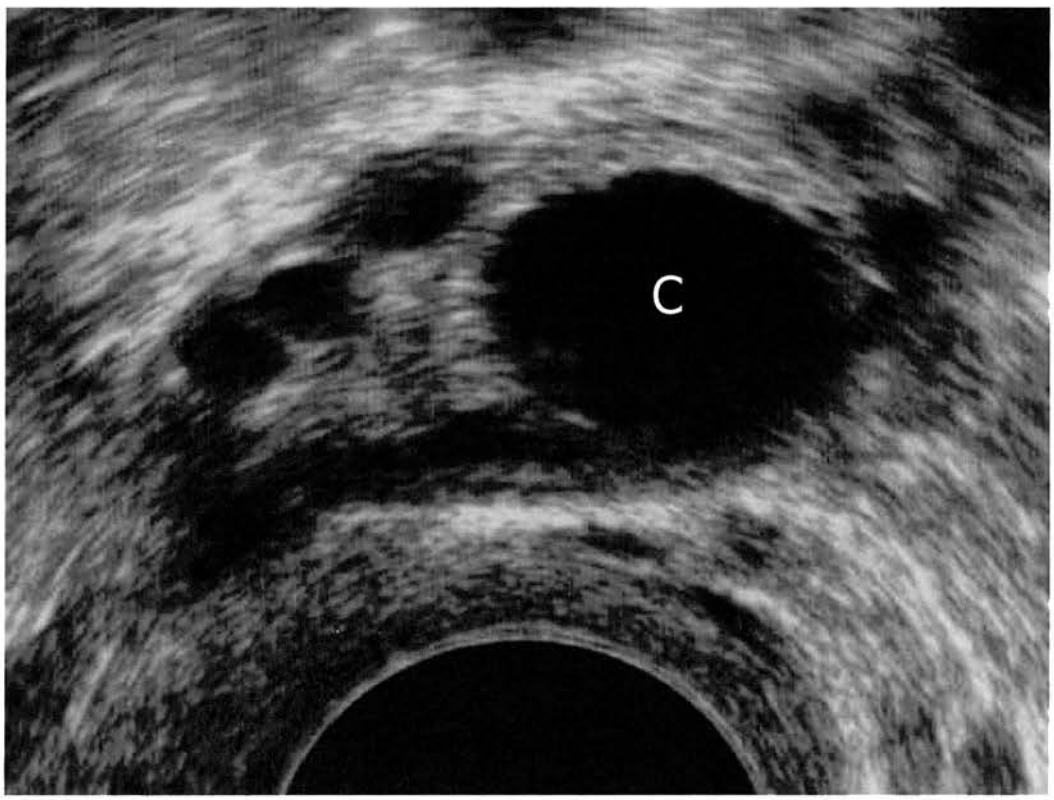
Ultrasound was first used to examine the seabed and was employed in 1912 to search for the wreck of the Titanic. This technique was fast tracked in the First World War to detect submarines and was known as SONAR (Sound Navigation and Ranging). In 1954, Ian Donald, then Professor of Midwifery at Glasgow University, was the first person to use ultrasound in obstetrics and gynaecology. He successfully differentiated a fibroid uterus from an ovarian cyst (Donald, 1958). At this time, ultrasound machines required immersion of the patient in a water bath to allow the use of water as a transmission medium. Donald built the first contact scanner using a jelly to achieve acoustic coupling between the skin and the transducer. Scientists at the Aloka plant in Japan added Doppler facilities and demonstrated blood flow in tissues using ultrasound (Kaneko, 1961). In the late 1970's miniaturisation technology enabled smaller transducers to be manufactured with the marketing of rectal and transvaginal probes. These were able to use

higher frequency ultrasound to achieve highly detailed images as the waves no longer needed to penetrate through the abdominal wall. In 1983, colour Doppler was developed enabling measurement of blood flow velocity and spectral analysis.

1.4.4 Morphology

Adnexal masses vary considerably in their structure, which in many cases reflects their histological derivation. Inspection of an adnexal mass with gray scale ultrasound gives a large amount of information about the origin and the structure of the mass. A number of investigators have assessed masses ultrasonically and attempted to classify them as either benign or malignant. Identification of the anatomical origin of the mass allows differentiation between ovarian, tubal and uterine masses and those arising from non-gynaecological organs (Figures 1.4 and 1.5).

Figure 1.4: A cystic corpus luteum within the ovary. The ovarian contour is smooth and encompasses the cyst (C denotes cyst). The cyst is unilocular and contains anechoic fluid.



The dimensions of the mass can be measured and the volume calculated. The internal structure of a mass can aid classification. A mass may be unilocular (comprising a single pocket of fluid); multilocular, containing a number of pockets, or it may be solid (Figures 1.4, 1.6 and 1.7).

Figure 1.5: A fimbrial cyst lying lateral to the ovary. The ovarian contour curves away from the cyst (C), which is clearly delineated by a rim of fluid beneath (arrowed). The cyst is again unilocular, containing anechoic fluid.

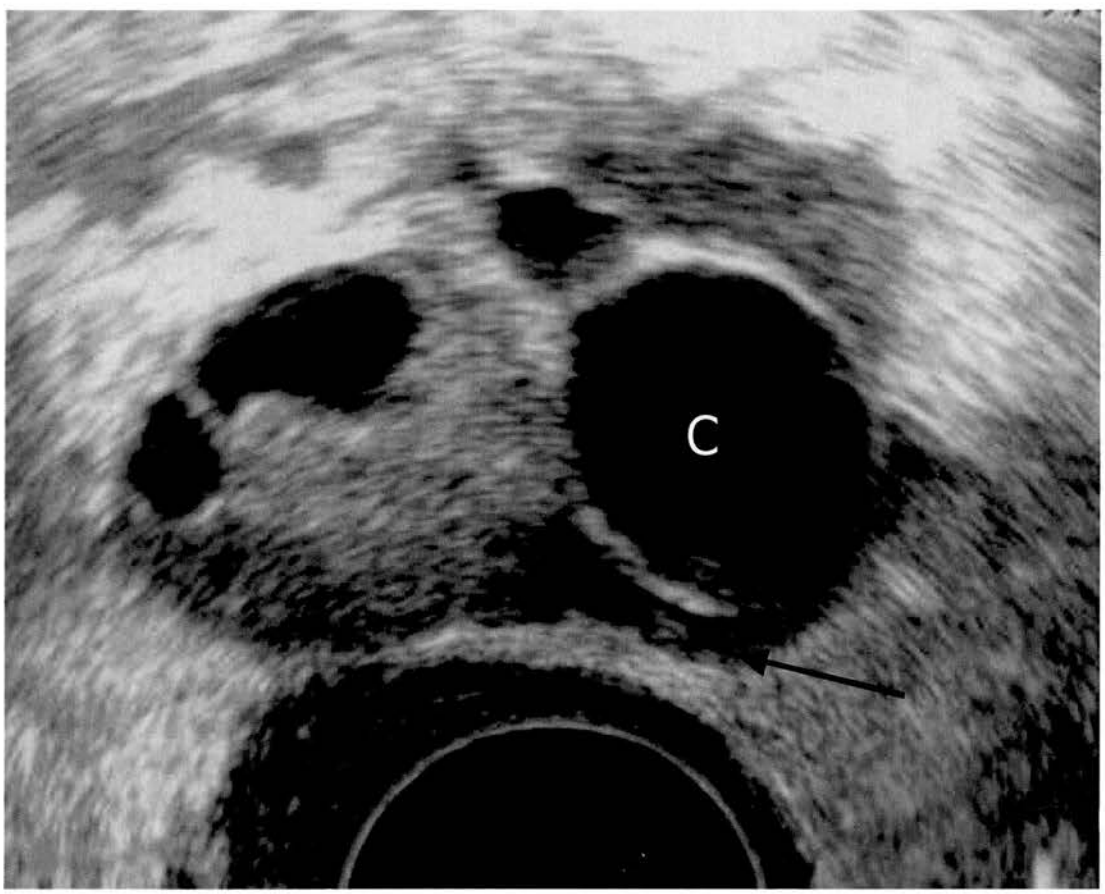


Figure 1.6: A mucinous cystadenoma. Note the characteristic multilocular appearance containing mucin of low-level echogenicity.

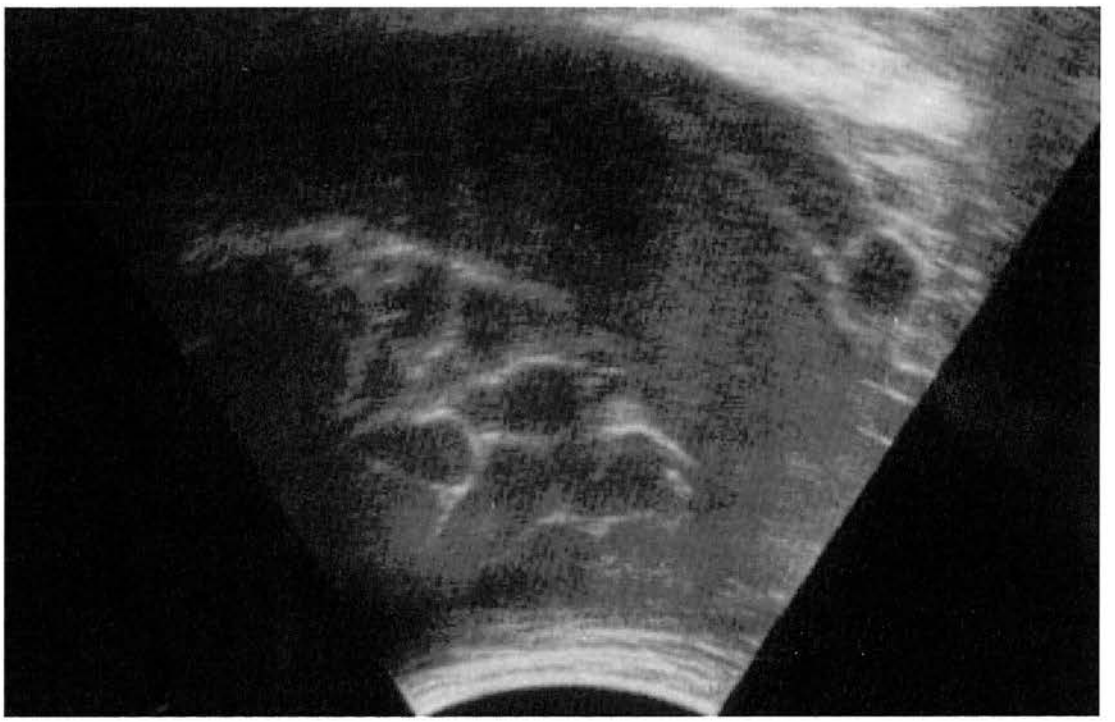
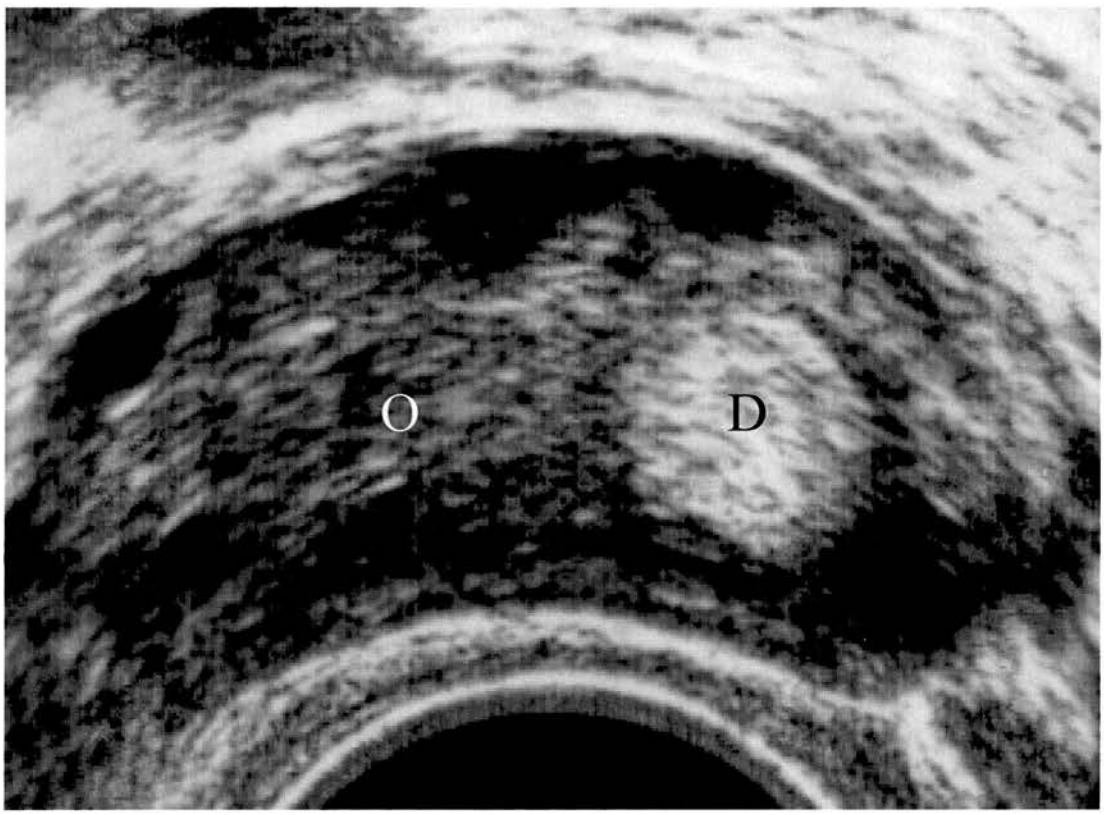


Figure 1.7: A small solid dermoid (D) placed laterally in the ovary (O) surrounded by healthy ovarian tissue.

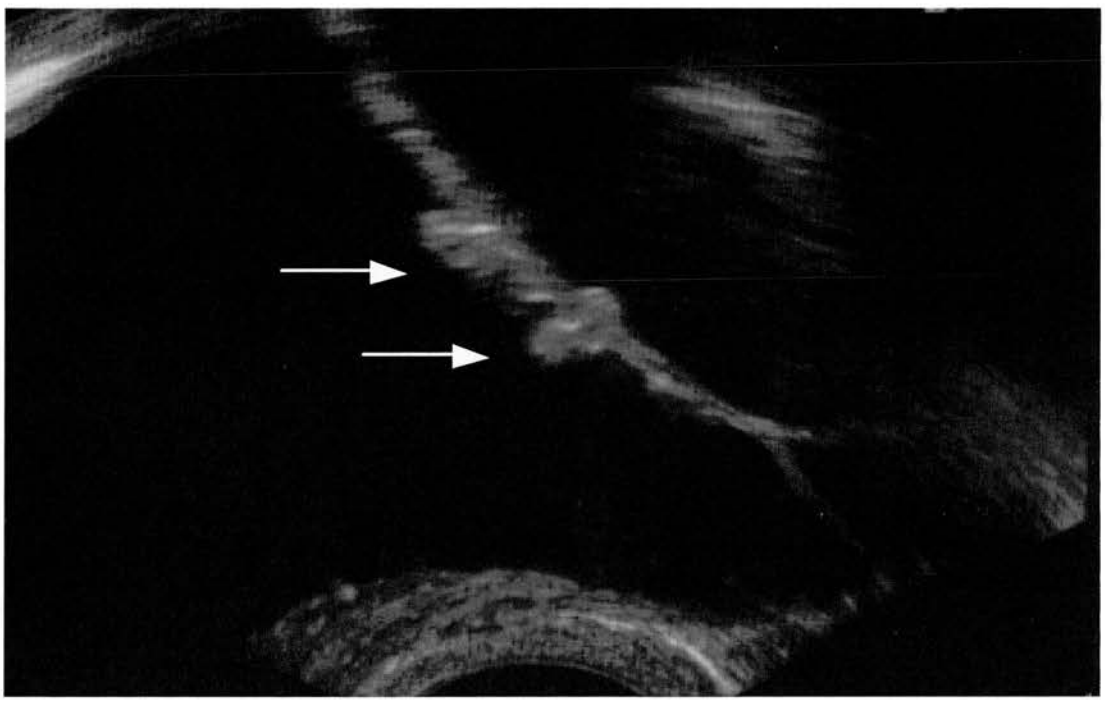


A series of 406 ovarian tumours showed that the risk of ovarian cancer was typified by increasingly complex and bizarre internal cyst structure. 76% of the ovarian cancers were correctly identified but a large number of dermoid cysts were mistakenly identified as malignant (Kobayshi, 1976). Eight morphological features were described that differentiated between benign and malignant cysts in 91% of cases (Meire et al, 1978). These were the size of the cyst, thickness of septae, presence of nodules, invasion of capsule and fixation of the cyst. Although the study involved examination of 184 recorded images, histological outcomes only were available for 51. A more robust study investigated the relation between the macroscopic appearance

of ovarian tumours at pathological examination and histological type. Of 296 unilocular cysts without internal or external vegetations, only one was malignant (0.3%) whereas 31 of 80 solid tumours were malignant (Granberg et al, 1989).

Ovarian volume was used to screen 6470 asymptomatic women for ovarian cancer. They were either postmenopausal or over the age of 30 with a family history of ovarian cancer. An abnormal scan was defined as an ovarian volume of greater than $10cm^3$ if postmenopausal; $20cm^3$ if premenopausal; or the presence of a papillary or complex tissue projection into a tumour (Figure 1.8). In 90 women, (1.4%) the scan was persistently abnormal. Eighty-two women had a laparotomy and eight, a laparoscopic oophorectomy. Of these, six women had primary ovarian cancer, one had metastatic colonic cancer, 77 had a benign tumour and six had a functional cyst. One screen negative woman developed a Stage IIA cancer 11 months after a normal scan. Although this gave an impressive negative predictive value of 0.9998, the positive predictive value was only 0.0667 (DePriest et al, 1997).

Figure 1.8: A serous cystadenofibroma. It is a multilocular anechoic cyst divided by septations bearing papillary projections (arrowed).



Granberg and co-workers scanned 180 women preoperatively and related the ultrasound features to the pathology of the adnexal mass (Granberg et al, 1990). They concluded that a number of morphological features increased the risk of malignancy: the presence of papillary proliferations; echogenic areas within the cyst and septations of the cyst. A papillary proliferation is a solid irregularity of the internal cyst wall which projects into the cyst cavity and may be smooth or irregular in contour (Figure 1.8). A strand of tissue dividing a cyst cavity into two locules is termed a septation. Using these features, they were able to differentiate between malignant and benign tumours with a sensitivity of 83% and a specificity of 92%.

This research led to the development of morphological scoring systems by a number of groups. The first scoring system was based on four morphological features: cyst wall thickness, inner cyst wall structure, septal thickness and echogenicity (Sassone, 1991). Each factor was scored, the total added and, if greater than nine, a preoperative diagnosis of malignancy made. The model was developed on a group of 143 women and achieved 100% sensitivity and 83% specificity with positive and negative predictive values of 37% and 100%, respectively. This model misdiagnosed a large proportion of dermoid cysts as ovarian malignancies; a similar finding to that of previous investigators (Kobayashi, 1976).

This system was refined with the use of multiple linear regression analysis. The wall thickness parameter was removed from the equation and the presence of shadowing inserted in an attempt to decrease the misdiagnosis of dermoid cysts. Each parameter was weighted to reflect its importance. On a larger study population of 312 women, a sensitivity of 96.8% and a specificity of 77% were achieved. The positive and negative predictive values were 29.4% and 99.6%, respectively. This poorer performance of the model was due to a persistently high number of false positive results despite the modification of the model (Lerner et al, 1993).

This model was tested prospectively in 2000 on a study population of 173 women. A sensitivity of 92% with a positive predictive value of 33% and a negative predictive value of 97% was achieved (Valentin, 2000). Although the results were reproducible in this study, other groups have found the model does not perform as well prospectively as in the original patient group (Ferrazzi et al, 1997, Mol et al, 2001). This may be due to a number of

factors; chief among them is the lack of definition of morphological parameters in the literature.

In an attempt to cross this barrier, the International Ovarian Tumour Analysis (IOTA) group led by Timmerman published a set of terms to describe the sonographic features of adnexal masses in 2000. The leading gynaecological scanners in Europe developed these terms. This enabled different groups to standardise the way they described ovarian cysts and allowed collaboration in a prospective multicentre trial. The IOTA trial reported in 2006 describing the different morphological criteria between borderline, primary invasive epithelial cancers, rare ovarian tumours and metastases to the ovary. They found that borderline tumours and Stage I primary epithelial invasive malignancies were commonly large, multilocular and contained papillary projections. Stage II-IV invasive epithelial malignancies were mostly solid with associated ascites. Metastatic tumours and rare ovarian tumours were generally solid and richly vascularised (Valentin et al, 2006).

1.4.5 IOTA classification of morphology

The adnexal lesion is defined as the part of an ovary or adnexal mass that is judged by ultrasonography to be inconsistent with normal physiology. If the lesion is a unilocular cyst within the ovary, surrounded by normal ovarian stroma, then the whole ovary containing the cyst is termed the 'ovary', and the cyst, the 'lesion'. If the abnormality is separate to the ovary e.g. a hydrosalpinx, then the ovary and lesion are measured separately. Both lesion and ovary are measured in three perpendicular planes. The volume of

the lesion is calculated from these measurements according to the formula for a prolate ellipsoid ($\pi/6$ x height x length x depth). The tumour origin is classified as ovarian, tubal, other or uncertain.

The presence of ascites is defined as fluid outside the pouch of Douglas. The depth of fluid in the pouch should be measured in its anteroposterior diameter in the sagittal plane.

A septum is defined as a thin strand of tissue running across the cyst cavity from one internal surface to the contralateral side. The thickness of the septum is measured where it appears widest (other than at its interface with the internal surface of the cyst wall). The operator should attempt to measure the septum at right angles to the ultrasound beam in order to decrease artefact and improve the accuracy of measurement.

An incomplete septum (as seen in a hydrosalpinx) is defined as a thin strand of tissue running across the cyst cavity from one internal surface to the contralateral side, but which is not complete in some scanning planes. If a cyst has only incomplete septae, it is unilocular, despite the fact that in some scanning planes it appears to be multilocular.

"Solid" is defined as echogenicity suggesting the presence of tissue e.g. myometrium or ovarian stroma. There are two methods to distinguish between solid tissue and a blood clot. The use of colour Doppler to identify flow within the tissue confirms solid tissue. The absence of flow is unhelpful. The second method is to gently push the structure with the transducer and look for internal movement suggestive of a clot. In doubtful cases, it should be classified as solid.

Solid papillary projections are defined as any solid projections into the cyst cavity from the cyst wall greater than or equal to 3mm in height. This arbitrary measurement was defined by the IOTA group and aims to exclude an irregular cyst wall but still include any significant solid projections into the cyst. The largest is measured in three planes: height, base and base. If it is unclear whether solid papillary projections or an incomplete septum is present, the worst-case scenario is used and the solid part should be classified as a papillary projection if their height is greater or equal to 3mm. The "white ball" in a dermoid should not be classified as a papillary projection.

The lesion is classified qualitatively into one of six classifications: unilocular; unilocular with a solid component; multilocular; multilocular with a solid component, or solid. A unilocular cyst is defined as a cyst without septae and with no solid parts or papillary projections. Normal ovarian stroma is not regarded as solid so a peritoneal pseudocyst encapsulating a normal ovary would be classified as unilocular rather than unilocular-solid. A unilocular cyst with a solid component is a cyst containing a measurable solid component such as a papillary projection. A multilocular cyst is one with at least one septum but no measurable solid component or papillary projections. A multilocular cyst with a solid component is one with septae and either a measurable solid component or at least one papillary structure. A solid tumour is one when at least 80% of its volume is solid. It may contain papillary projections protruding into the small cystic spaces within the tumour. Some dermoids display a 'tip of the iceberg' sign or have contents that are unclear: these are termed 'not classifiable'.

The true number of locules within the lesion is noted after assessment in various planes. In cystic solid tumours, the largest solid area is measured in three different planes. In some cases, the largest solid component may be a papillary projection; this should be measured as both the largest solid component and largest papillary projection.

The internal wall is noted as being smooth or irregular. If there is a solid papillary projection then the wall is irregular by definition. The external wall of the lesion is not inspected. If the tumour is solid, the internal wall cannot usually be classified as smooth or irregular but the external border of the tumour is described as smooth or irregular. If there is any irregularity in either the inner wall of any cyst or the surface of a solid component, the lesion is described as irregular.

The dominant feature of the cystic contents is described as anechoic (black); low-level echogenic (homogeneous low level echogenicity as seen in mucinous tumours); ground glass (homogeneously dispersed echogenic cystic contents as seen in endometriotic cysts); haemorrhagic (with internal thread-like structures, representing fibrin strands, which may appear star-shaped, cobweb-like or jelly-like) or mixed echogenicity (as often seen in dermoid cysts). In a solid tumour, the dominant feature of any cystic contents is described only if it can be assessed.

The presence of acoustic shadows, defined as loss of acoustic echo behind a sound reflecting structure, is also noted.

1.4.6 Subjective assessment

Some tumours have a distinct ultrasonographic appearance allowing recognition of their histological type. With experience, an ultrasonographer can often make a presumptive diagnosis based on the gray scale appearance of the tumour alone. For example, a dermoid cyst is characteristically located laterally in the ovary, surrounded by a rim of normal ovarian tissue (Figure 1.7). They typically display mixed echogenicity and hair inside the cyst may be recognised by the presence of spiculations. Areas of calcification casting acoustic shadows may be seen due to bone or tooth formation (Figure 1.2). 173 women were scanned preoperatively and in 42% of cases, a correct specific diagnosis was made based on the gray scale image alone e.g. endometrioma or dermoid cyst. An incorrect specific diagnosis was made in 7% and in 51%, a specific diagnosis was not reached. The diagnosis was then reconsidered following the acquisition and analysis of the Doppler waveform. In 3%, it changed the diagnosis from an incorrect one to the correct specific diagnosis and in 1% it changed from correct to incorrect (Valentin, 1999a).

Subjective assessment is a reliable tool for the diagnosis of histological type in those tumours displaying a typical morphology. However, in over half the cases a specific diagnosis was not reached. The value of additional Doppler interrogation was limited (Valentin 1999a).

1.4.7 Doppler assessment

The growth of malignant tumours is dependent on the diffusion of oxygen and nutrients from their surroundings. A solid tumour cannot grow beyond

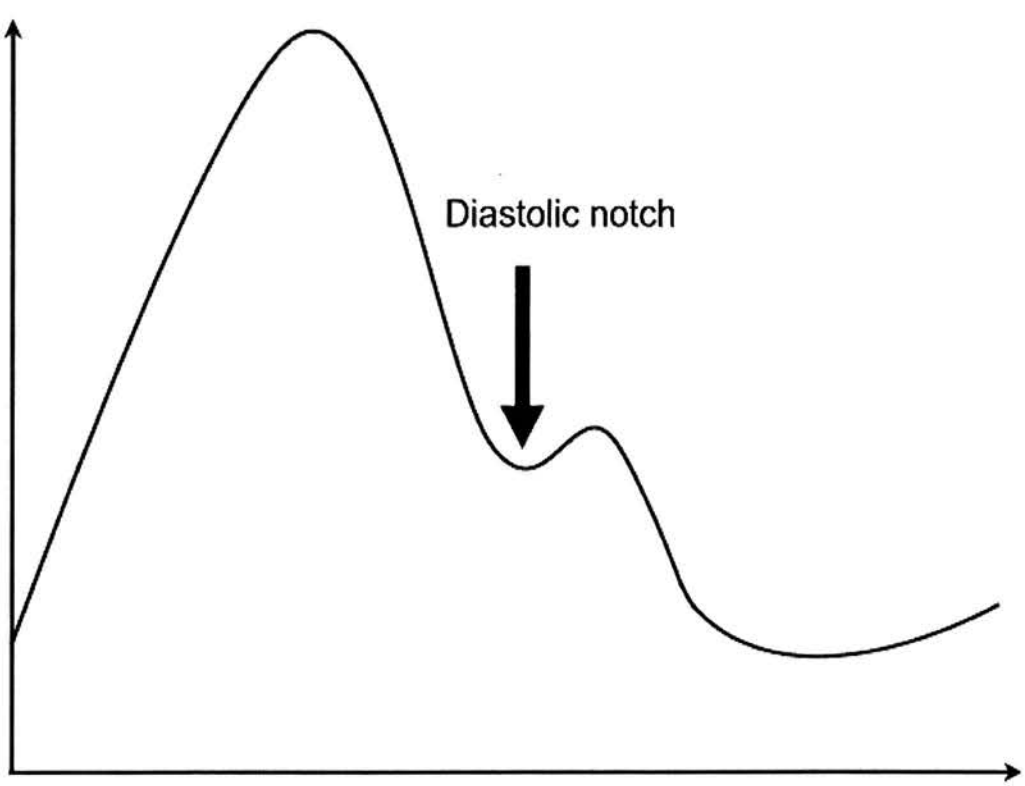
one square millimetre without neovascularisation (Folkman, 2003). Tumour angiogenesis involves migration and proliferation of endothelial cells and is stimulated by cytokines including fibroblast growth factors, angiogenin and vascular endothelial growth factor. Both cancer cells and other surrounding cells produce these cytokines. They stimulate the development of new arcades of vessels that penetrate the tumour from its external borders. The new vessels display both structural and functional differences from normal vessels. They are relatively abundant and display an irregular branching pattern with arterio-arterial anastomoses; arterio-venous shunts and, occasionally, thin walled vessels ending in tumour lakes. The functional differences include an incomplete elastic smooth muscle layer leading to poor control of distal perfusion. This results in little resistance to flow and decreased variation in flow through the cardiac cycle. Poor endothelial function renders the vessel more permeable to large molecular weight proteins that increase the oncotic pressure surrounding the vessels. Arterio-venous shunts result in high velocity flow due to the large pressure gradient. These changes lead to a wide variation in vascular flow within a tumour (Vaupel, 1994).

Colour Doppler imaging can be used to visualise the structural changes in vascular architecture and the functional changes can be assessed with spectral Doppler. The lack of vasomotor control and the presence of arterio-venous shunts produce vessels demonstrating rapid flow with low resistance. This gives rise to high peak systolic velocities with low resistance and pulsatility indices.

A Croatian group used Doppler imaging to differentiate benign from malignant adnexal masses (Kurjak, 1991). In the context of a screening

study for ovarian cancer, they identified 690 adnexal masses, 47 of which were malignant primary ovarian cancers. They found the RI to be below 0.40 for all the malignant masses. Other groups also found a clear discrimination between benign and malignant masses with high velocity, low resistance flow in cancers and little or no flow in benign masses (Bourne et al, 1989). Three parameters were described for the identification of benign ovarian masses: peripheral location of vessels; high impedance parameters, and a diastolic notch in the Doppler waveform (Figure 1.9). A diastolic notch is caused by the recoil of the elastic smooth muscle coat of the vessel wall following systole, which is characteristically absent in malignant vessels. Ovarian cancers demonstrated centrally placed vessels and an absence of the diastolic notch (Fleischer, 1993a).

Figure 1.9: A diastolic notch in the Doppler waveform.



Further work compared the specificity of both gray scale and colour Doppler imaging to identify the histology of adnexal masses. Doppler interrogation was found to be helpful in identifying ovarian malignancies, ovarian torsion and ectopic pregnancy. A greater accuracy was evident in postmenopausal rather than premenopausal women (Fleischer et al, 1993). This may have been due to the significant functional vascular changes in the normal ovary with the high degree of neovascularisation during the development of the corpus luteum. These vessels display a high peak systolic velocity and low resistance index (Jauniaux et al, 1992), similar to that described in ovarian malignancies. However, a significant overlap in Doppler signals from benign and malignant masses was seen, in contrast to previous studies (Fleischer et al, 1993).

Further doubt was shed on the use of Doppler in 1994. A wide range of PI was recorded in different vessels within a single mass and a considerable overlap was found between the range found in benign and malignant tumours. PSV and TAMXV appeared to be better discriminators than PI between benign and malignant cysts. A subsequent study found that Doppler interrogation of solid pelvic tumours showed no significant difference between TAMXV in benign and malignant tumours (Sladkevicius et al, 1995).

A combination of two parameters (TAMXV and PI) improved the detection of malignancy in a set of 51 women. The combined use of a cut off level of greater than 12cm/s for TAMXV and a PI less than or equal to 1.0 gave a sensitivity of 88.9% and a specificity of 88.1% (Tailor et al, 1996). Colour

Doppler energy (CDE) was used to assess intraovarian blood flow and compared to the findings of colour Doppler imaging (CDI). There was no significant difference in discrimination between the two methods (Tailor et al, 1998).

The intraobserver reproducibility of CDI was found to be greater than that of CDE (Tailor et al, 1996). This may be due to the aliasing effect, which identifies the areas with highest velocity flow and therefore is likely to cause consistent placing of the Doppler gate at these points.

A Japanese group found PSV and serum CA 125 to be the most discriminatory variables for the diagnosis of malignancy in 171 women with an adnexal mass. Hata found a relation between prognosis and peak systolic velocity in women with ovarian cancer. Worsening histological grade, a lower apoptotic index and higher microvessel count were significantly associated with a higher preoperative intratumoral PSV. However, he found a wide variation in PSV in the 49 women recruited with a range of 4.3 cm/s to 80.9 cm/s (Hata et al, 2002).

1.5 Serum tumour markers

A serum tumour maker is a biochemical compound that may be produced by, or released in response to, a tumour. They can be either tumour specific or tumour associated antigens. Tumour specific antigens are unusual and are typified by immunoglobulins released from B lymphocyte tumours. Tumour associated antigens make up the majority of tumour markers in clinical use. Although they are raised in the presence of a tumour, they are also produced both in physiological states and in benign and other malignant diseases. Most tumour-associated antigens are macromolecular and may be enzymes, hormones, receptors, growth factors, biological response modifiers or glycoconjugates.

1.5.1 CA 125

CA 125 is an antigen expressed by fetal amniotic and coelomic epithelium. In adults, it is found in tissues derived from coelomic epithelium such as the mesothelial cells of the pleura, pericardium and peritoneum. It is also found in the mullerian epithelia of the tube, endometrium and endocervix. Expression has also been discovered in epithelial cells outside the genital tract including lung, breast and conjunctiva (Nap, 1998). The surface epithelium of normal ovaries in the adult does not express CA 125 except in inclusion cysts, areas of metaplasia and papillary proliferations (Kabawat et al, 1983).

It was first detected in 1981 by Bast et al, using a murine monoclonal antibody OC125, raised in response to an immunologic challenge with an ovarian cancer cell line (Bast et al, 1981). It is now known to be a mucin-type glycoprotein with a molecular weight of over 200,000 daltons. It is highly glycosylated and is rich in the amino acids serine, threonine and proline. It has two major antigenic domains classified as A, the domain binding monoclonal antibody OC125; and B, a domain binding monoclonal antibody M11.

The original immunoradiometric assay was a homologous assay using the monoclonal antibody OC125 alone. This was prone to variability between different CA 125 kits, which gave discordant, and in some cases discrepant results. A second-generation heterologous CA 125 assay using both monoclonal antibodies (OC125 and M11) has since been developed. It has been shown to give better analytical performance with fewer false positive results at the same cut off value of 35 units per millilitre (U/ml) (Kenemans et al, 1993). CA 125 was shown to be below 35 U/ml in 99% of healthy female donors (Bast et al, 1983). This has been accepted as the upper limit of normal in the general population. CA 125 levels are influenced by both physiological and pathological events. Menstruation is associated with a modest rise in CA 125 in 30% of women (Alagoz et al, 1994). CA 125 levels are raised in the first trimester of pregnancy, with a median serum CA 125 level of 23.4 U/mL and a 95% reference interval of 5.28-70.15 (Aslam et al, 2000a). Benign conditions such as endometriosis, uterine fibroids, tubo-ovarian abscess, benign ovarian cysts, ovarian hyperstimulation syndrome and diverticulitis can all cause a rise in CA 125 titres (Ozaksit et al, 1995). As the bulk of benign causes of a raised CA 125 level are seen only in premenopausal women, some authors have advocated a lower cut off level

for postmenopausal and hysterectomised women of 20 U/ml. (Alagoz et al, 1994). Malignant processes in the ovary, breast, colon, bladder, pancreas, liver and lung are associated with an elevation of serum CA 125 levels.

Epithelial ovarian cancer is associated with an elevated CA 125 level in 85% of patients (Bast et al, 1983, Canney et al, 1984). Over 90% of women with advanced stage ovarian cancers have raised levels whereas only 50% of women with Stage I cancer have an elevated level. Different histological types of ovarian cancer are associated with varying levels of CA 125 elevation. Serous epithelial cancers are associated with the most significant rise in CA 125 levels whereas mucinous and clear cell tumours show more modest rises. Borderline tumours show a small elevation in CA 125 levels. The mean CA 125 in borderline tumours is 44kU/l with a level of 1201 U/ml in epithelial and 53 U/ml in non-epithelial ovarian tumours (Aslam, 2000a).

A raised CA 125 can be detected in asymptomatic women prior to the diagnosis of ovarian cancer. Zurawaski in 1988 analysed 59 serum samples from women who were diagnosed as having ovarian cancer 5 years later and found a raised level in 25%. Skates assayed serum samples from 3554 asymptomatic Swedish women who had been enrolled in a screening study, 6 of whom went on to develop ovarian cancer. He found that CA 125 levels demonstrated an exponential rise before the clinical diagnosis of ovarian cancer (Skates 2003). Einhorn investigated the use of CA 125 in the pre-operative discrimination of adnexal masses (Einhorn et al, 1986). Using a cut-off of 35 U/ml a raised CA 125 had a sensitivity of 78% and a specificity of 95% in the detection of malignancy.

1.5.2 CA 19-9

CA 19-9 is a mucin-type glycoprotein with a molecular weight of approximately ten kilodaltons. It is defined as the antigen reacting with the monoclonal antibody 1116 NS 19-9. This antibody was originally obtained by inoculating BALB/c (albino strain) mice with a cell line (SW 1116) derived from a human colorectal carcinoma.

The biological function of CA 19-9 is not clear. It has been suggested that it may play an important role in metastasis, perhaps by acting as an anti-adhesion molecule. CA 19-9 is elevated in ovarian mucinous cystadencarcinoma and in other primary malignancies of the pancreas, gastrointestinal tract, lung, bile duct and endometrium. It is expressed on cell surfaces, particularly biliary and pancreatic duct cells and is raised in 80% of cases of pancreatic cancer and 70% of biliary cancers (Maestranzi et al, 1998). Despite the identification of other mucin markers for pancreatic cancer, CA 19-9 has been established as the 'gold standard' against which other markers are evaluated.

CA 19-9 is also elevated in metastatic disease of the liver. Benign diseases that may cause elevated CA 19-9 include pancreatitis, hepatocellular jaundice and hepatic cirrhosis. It is significantly raised in obstructive jaundice caused by stones in the common bile duct (Duffy 1998). CA 19-9 is thought to be cleared by hepatic metabolism and excreted in bile, which would explain the high concentrations in extra-hepatic biliary obstruction (Maestranzi et al, 1998).

Human mucinous epithelial ovarian cancer cell lines have been found to secrete CA 19-9 into their culture medium (Sato et al, 2002). 55.9% of women

with epithelial ovarian cancer have an elevated serum CA 19-9 level (Kudoh et al, 1999). The incidence of a raised CA 19-9 is higher in mucinous carcinomas (76.9%) but is unrelated to tumour stage. All women with a serum ratio of CA 125 to CA 19-9 of greater than 50 have either serous or endometrioid cancer (Kudoh et al, 1999). Contrary to this, levels of CA 19-9 in Stage Ic ovarian cancer were found to be significantly higher than those of Stage Ia cancers (565 U/ml vs 26 U/ml) (Murumatsu et al, 2005). CA 19-9 can be used to discriminate between malignant and benign ovarian masses. A cut-off of 40 U/ml gave a specificity of 81.1% with a sensitivity of 35.6% with levels raised more commonly in malignant mucinous tumours (Gadducci et al, 1992). CA 19-9 is also raised in 46% of borderline ovarian tumours and in 57% of mucinous borderline tumours. Only 15% of CA125 levels were raised in the mucinous group (Engelen et al, 2000).

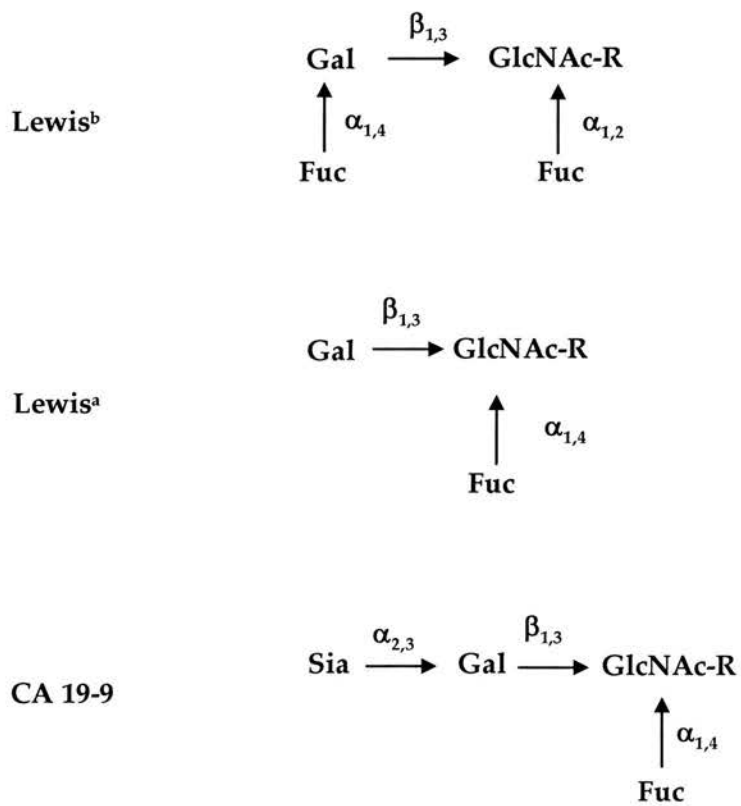
Tissue expression of CA19-9 has been investigated in an attempt to identify tumour markers that would complement CA 125. CA 19-9 is expressed in 29% of ovarian tumours which do not express CA 125. However, only 2% of the tumours they studied were mucinous which may have led to an underestimation of the number of tumours which express CA19-9 (Rosen et al, 2005). CA 19-9 is elevated in the pressence of dermoid cysts with a mean of 101 U/ml. CA 125, carcinoembryonic antigen (CEA) and alpha fetoprotein (AFP) are not raised with dermoids. A raised CA 19-9 is significantly associated with bilateral dermoid cysts (Dede et al, 2006).

Carbohydrate antigens are synthesised along the same pathway as normal human blood group antigens. The Lewis blood group antigens Lewis a and Lewis b are carbohydrate structures that form epitopes on glycolipids and glycoproteins. The CA 19-9 level in the serum of an individual is dependent

on the blood group antigens expressed by the individual. The CA 19-9 antigen is a sialylated form of the Lewis a (Lewis^a) antigen. The Lewis gene (Le) encodes a fucosyltransferase enzyme (Fuc) that is responsible for adding fucose in an $\alpha_{1,4}$ linkage to the N-acetylglucosamine (GlcNAc-R) of blood group oligosaccharides (Figure 1.10) (Itzkowitz, 1986).

Figure 1.10: The synthesis pathway of the Lewis a antigen, the Lewis b antigen and CA 19-9.

The Lewis b antigen is formed by the addition of two fucose molecules (Fuc) to N-acetylglucosamine (GlcNAc-R) with the use of two fucosyltransferase enzymes and then the addition of a galactose (Gal) molecule by the galactosyltransferase enzyme. The Lewis a antigen is formed by the action of fucosyltransferase enzyme and the galactosyltransferase enzyme adding fucose (Fuc) and galactose (Gal) respectively. The CA19-9 antigen is formed by the sialation of N-acetylglucosamine (GlcNAc-R) by a $\alpha_{2,3}$ sialyltransferase enzyme (Sia). The galactosyltransferase enzyme then adds galactose (Gal) and the fucosyltransferase enzyme adds fucose (Fuc) with an $\alpha_{1,4}$ linkage to form the CA19-9 antigen.



Because around 7% of the population lacks the Le gene, they are unable to synthesise the CA 19-9 antigen. As a result, Lewis negative individuals with

the blood type Lewis (a-b-) will have undetectable concentrations of CA 19-9 whatever the tumour burden. Clinically therefore, the maximal achievable sensitivity of the CA 19-9 assay cannot be greater than 93% (Vestergaard 1999).

1.5.3 CA 72-4

Cancer antigen 72-4 is a high molecular weight mucin that is found in many carcinomas, particularly in colonic, gastric and ovarian primaries. It is also known as tumour associated glycoprotein 72 (TAG 72). One of its epitopes, sialosyl-2-6α –N-acetylgalactosaminyl, is recognised by the monoclonal antibody B72.3. Serum contribution of CA 72-4 is minimal as the antigen is only expressed in a few normal tissue types including secretory endometrium and transitional colonic mucosa.

CA 72-4 has been shown to be elevated in 40% of gastrointestinal adenocarcinomas. A combination of CA 72-4 and carcinoembryonic antigen (CEA) has been used for the detection of colorectal cancer and in post-surgical follow up. In patients with no clinical evidence of disease recurrence, both CEA and CA 72-4 remained negative (Guadagni et al, 1992).

The normal range of CA 72-4 is 0 to 6 U/ml. CA 72-4 is elevated in patients with both ovarian and gastrointestinal cancer, with a lower degree of expression in patients with breast, prostate and lung cancer. It is rarely elevated in patients with benign disease (Gero, 1989).

Immunohistochemical analysis of ovarian cancer has shown CA 72-4 antigen to be present in the majority of histological types of cancer (Thor et al, 1986). Serum levels are raised in around 50% of patients with ovarian cancer. The assay has a low percentage of false positives and therefore has been suggested as a confirmatory marker in women with a high CA 125 (Gero, 1989).

The sensitivity of CA 125 has been compared to CA 72-4 in the diagnosis of ovarian cancer. Both markers had a similar sensitivity for the detection of ovarian cancer but CA 72-4 had a better sensitivity in mucinous cancer and CA 125 in serous cancer. A combination of the two markers gave no advantage for follow-up (Hasholmer et al, 1996).

Eight tumour markers were evaluated in women with a pelvic mass. At a cut off level for CA 72-4 of 3.8 U/ml, a sensitivity of 54.2% and a specificity of 91.6% was achieved. A logistic regression model combined CA 125, OVX1, CA 15-3 and lipid associated sialic acid (LASA), which gave a sensitivity of 82.3% and a specificity of 93.2% in the preoperative detection of ovarian cancer (Woolas et al, 1995).

50.5% of women with ovarian cancer have a level of CA 72-4 over 4 U/ml. The performance of CA 72-4 in the prediction of ovarian cancer is 69.1%, measured by the area under the curve (AUC) of a receiver-operator characteristic curve. This is similar to the performance of CA 125 which gives an AUC of 69.0% (Udagawa et al, 1998).

An Italian group measured the serum levels of CA 72-4 in 726 women. They found that 66% of women with ovarian cancer showed elevated CA 72-4 levels and that these correlated to the clinical course of the disease through

chemotherapy and follow-up. However, they did not find that addition of CA 72-4 to CA 125 increased sensitivity (Scambia et al, 1990).

A Dutch group investigated the addition of CA 72-4 measurement after physical examination, ultrasound examination and CA 125 estimation. They developed a logistic model in a series of 155 women with a pelvic mass. CA 72-4 alone gave a sensitivity of 61% with a specificity of 93% and an odds ratio of 4.9 in the prediction of malignancy. The addition of CA 72-4 to the logistic model improved the diagnostic accuracy from 81 to 87% (Schutter et al, 1997).

Skates et al investigated the use of CA 72-4 in combination with CA 125 and macrophage colony stimulating factor (M-CSF) in screening for early stage ovarian cancer. With the use of logistic regression, classification trees and mixture discriminant analysis, they were able to detect early stage cancer with a sensitivity of 70% and a specificity of 98% (Skates et al, 2004).

1.5.4 CA 15-3

CA 15-3 is a heterogeneous high molecular weight mucin, which is found in the human milkfat globule membrane. It is coded for by the MUC1 gene and produced by secretory epithelium. It is over-expressed in breast, ovarian and lung cancers and shed into the serum. CA 15-3 is recognised by two antibodies, DF3 and 115D8. The DF3 antibody was prepared using a membrane-enriched fraction of a human breast cancer cell line and the 115D8 antibody was raised against antigens found in the human milkfat globule membrane (Cheli et al, 1998).

CA 15-3 levels are elevated in breast, ovarian and colorectal carcinoma (Walach et al, 1998). CA 15-3 has been shown to be a useful marker in metastatic breast cancer and indicates both disease progression and response to therapy. It has also been found to rise prior to clinical detection of disease recurrence (Safi et al, 1991, Cheli et al, 1998).

Drapkin et al in Boston, USA, studied MUC1 expression in papillary serous and endometrioid ovarian carcinomas and compared it to inclusion cysts and normal ovarian epithelium. They found that the ovarian cancers were immunopositive for MUC1 whereas both the inclusion cysts and normal ovarian epithelium were negative for MUC1 (Drapkin et al, 2004). Work by Feng and colleagues showed that 90% of serous tumours, 62% of mucinous tumours and 80% of clear cell tumours expressed MUC1. They found that over expression of MUC1 was associated with a higher histological grade and stage than those tumours that did not over- express MUC1. They concluded that overexpression of MUC1 was linked to a worse prognosis in ovarian cancer (Feng et al, 2002).

A combination of CA 15-3 and CA 72-4 was used to differentiate between sera from women with elevated CA 125 with benign disease and ovarian cancer. At least one of CA 15-3 or CA 72-4 was elevated in 77% of the cancers but only in 6% of the benign cases. These markers could therefore be used to increase the specificity of detection of ovarian cancer in elevated levels of CA 125 (Bast et al, 1991). The use of CA 15-3 for the discrimination between benign and malignant ovarian masses gave a sensitivity of 57.1% with a specificity of 93.9% (Gadducci et al, 1992).

Woolas found CA 15-3 to give a sensitivity of 62.5% with a specificity of 86.5% at a cut-off level of 32 U/ml in a population of women with pelvic masses. He combined this value with four other markers as described in section 1.5.3 (Woolas 1995).

1.5.5 HER-2/neu

HER-2/neu is a human oncogene that encodes a transmembrane growth factor receptor with a molecular weight of 185 000 daltons. It is also known as c-erbB-2 and p105. The protein is made up of three portions: an internal cytoplasmic structure with tyrosine kinase activity; a short hydrophobic transmembrane section and an extracellular ligand-binding domain (ECD). The ECD has a 44% sequence homology with the human epidermal growth factor receptor. It has a molecular weight between 95 000 and 115 000 daltons. The ECD is shed into the bloodstream and can be measured in serum.

The level of shed ECD is elevated in women with breast cancer, particularly in the presence of metastatic disease. The HER-2/neu gene has been found to be amplified in 25 to 30% of human breast cancers. Breast cancer patients with multiple copies of the HER-2/neu gene have both a shorter time to relapse and shorter overall survival. Multivariate analysis found HER-2/neu HER-2/neu gene amplification to be the second strongest predictor after lymph node positivity (Slamon et al, 1989). Raised serum concentrations of HER-2/neu are associated with tumours that are more aggressive.

HER-2/neu is also expressed by most ovarian cancers and is over-expressed in approximately 30%. Slamon et al showed that gene amplification in ovarian carcinoma was associated with decreased survival. Tumours with a single copy of the HER-2/neu gene conferred a life expectancy five times longer than those with more than five copies (Slamon et al, 1989). Tissue expression of HER-2/new was investigated in frozen sections of normal ovarian tissue and advanced epithelial ovarian cancer (Berchuk et al, 1990). Thirty-two percent of the epithelial cancers stained heavily for HER-2/neu. The survival of these patients was significantly worse than those whose tumours did not over-express HER-2/neu: 15.7 months vs 32.8 months respectively (p=0.001).

McKenzie et al investigated the relationship between tumour over-expression and serum levels of HER-2/neu (McKenzie et al, 1993). They found that the serum level was elevated in 15% of women with ovarian cancer whereas 6% of healthy volunteers had elevated levels. Immunohistochemistry of the ovarian cancer tissue showed 38% had over expression of HER-2/*neu*. The serum *neu* levels had a sensitivity of 29% with a specificity of 93% in the determination of tumour over expression of HER-2/neu. However, half of the serum samples were obtained post-operatively and the authors postulated that the levels might have dropped following decrease in the tumour burden.

Meden and co-workers established a normal range for HER-2/neu and demonstrated increasing levels during pregnancy, which peaked in the third trimester (Meden et al, 1994). In a later publication, they related an elevated serum level of HER-2/neu to a poor prognosis in ovarian cancer. They used a cut-off of two standard deviations above the normal mean. Women with a

raised serum HER-2/neu had a survival of 7 months as opposed to 29 months. They postulated that a raised serum HER-2/neu defined a subgroup of particularly aggressive tumours (Meden et al, 1997).

In vitro work demonstrated that over-expression of HER-2/neu in a breast cancer cell line induces endothelial cell retraction, enhancing the metastatic potential of the tumour cells. Herceptin® (a recombinant anti-HER2 antibody) significantly blocked this endothelial cell retraction (Carter et al, 2001).

50% of borderline ovarian tumours express HER-2/neu (Haaften-Day, 1996). Eltabbakh and co-workers related HER-2/neu over-expression in borderline tumours to FIGO staging, recurrence and survival. They found that of 42 tumours, 21% demonstrated over-expression of HER-2/neu. Over-expression was significantly more common in stage III than stage I tumours (6/12 vs 3/30, $p = 0.0157$). No relationship between optimal debulking and over-expression was identified and there were no recurrences in the follow-up period (Eltabbakh et al, 1996).

1.6 Statistical models

1.6.1 Risk of malignancy indices

Based on the success in discriminating between benign and malignant ovarian masses using morphological markers, they were combined with demographic variables and tumour markers to produce a risk of malignancy index (RMI). This model multiplied the CA 125 level (U/ml) by the menopausal status (1 for premenopausal or 3 for menopausal) and an ultrasound score (Jacobs et al, 1990).

RMI = CA 125 x menopausal score x ultrasound score

The presence of multilocular cysts, bilateral lesions, solid areas, ascites and intra-abdominal metastases scored 1 point each. The total ultrasound score was calculated with a score of 0 for 0 points, 1 for 1 point and 3 for 2 or more. The most discriminatory cut-off value was 200 for the diagnosis of malignancy (Table 1.3). On a population of 143 patients (101 benign, 42 malignant), a sensitivity of 85% with a specificity of 97% was achieved. This out-performed CA 125, the ultrasound score, age, menopausal status and the clinical impression of malignancy.

This model is the best-known formula for the discrimination between benign and malignant cysts. It is user-friendly as the ultrasound marking score can be calculated with reference to an ultrasound report and the overall score requires straightforward mental arithmetic. It can be used in any

department with facilities for CA 125 estimation and basic ultrasound evaluation.

Jacobs' model was refined using stepwise forward logistic regression to identify the most significant variables. CA 125 was entered as a continuous variable and menopausal status as one for premenopausal and four for postmenopausal women. The ultrasound score was scored on the above criteria but scores of 0 or 1 were given a value of 1 whereas scores of 2 and above were given a value of 4. On a Norwegian population, the refined RMI gave a sensitivity of 95% with a specificity of 87% whereas Jacobs' model gave a sensitivity of 87% and a specificity of 91% at a cut-off of 200. Jacobs' model performed less well in this study than in its original publication. The model was designed on retrospective data analysis giving a best fit to that cohort (Tingulstad et al, 1996).

Table 1.3: Sensitivity, specificity and likelihood ratio for malignancy given a positive or negative result for different levels of the risk of malignancy index (after Jacobs, 1990)

	Sensitivity		Specificity		Likelihood ratio for malignancy if result is	
RMI score	**%**	**95% CI**	**%**	**95% CI**	**Positive**	**Negative**
25	100.0	91.4 -100.0	62.2	51.9 – 71.8	2.7	0.00
50	95.1	83.5 – 99.4	76.5	66.9 – 84.5	4.1	0.06
75	92.7	80.1 – 98.5	84.7	76.0 – 91.2	6.1	0.09
100	85.4	70.8 – 94.4	87.8	79.6 – 93.5	7.0	0.17
150	85.4	70.8 – 94.4	93.9	87.2 – 97.7	14.0	0.16
200	85.4	70.8 – 94.4	96.9	91.3 – 99.4	42.1	0.15
250	78.0	62.4 – 89.4	99.0	94.5 – 100.0	76.9	0.22

1.6.2 Logistic regression models

Anil Tailor in London developed a logistic regression model in 1997 based on demographic and ultrasound criteria. He used logistic regression to identify the most important of ten variables (age, maximum tumour diameter, tumour volume, unilocularity, papillary projections, random echogenicity, highest PSV, TAMXV, PI and RI). The retained variables were age, papillary projection score (0 for absent, 1 for present) and time averaged maximum velocity (TAMXV).

The equation is shown below:

$$\textbf{Probability} = \frac{1}{(1 + e\text{-}z)}$$

Where:

$$\textbf{z = (0.1273 x age) + (0.2794 x TAMXV) + (4.4136 x papillary projection score) – 14.2046}$$

A cut off of 0.5 was used which gave a sensitivity of 81.8% with a specificity of 98.1%. A separate group of 15 cases was used to test the model; all were correctly classified. An overall sensitivity of 86.7% and specificity of 98.1% were achieved. A strength of this study was the use of a test set to prospectively validate the model. The use of stepwise logistic regression allowed independent evaluation of the variables with a calculated rather than arbitrary weighting. However, the numbers in the study were small with a total population of 67 women. Among them were 15 malignant cases with 12 invasive epithelial cancers and 3 borderline tumours (Tailor et al, 1997).

Alcazar and Jurado in Pamplona developed an LRM based on morphological criteria as described previously by Sassone and the vascularity of the tumour (resistance index greater than 0.45). They prospectively tested their model on 58 patients and achieved a sensitivity of 84.6% with an impressive specificity of 100% (Alcazar and Jurado, 1997).

Timmerman, Bourne and Tailor collaborated in 1999 to develop a further LRM (Timmerman et al, 1999a). They used a subset of their data to test it prospectively against Jacobs' RMI and Lerner's morphological scoring

system. Their model was the first regression model to combine menopausal status, CA 125 levels and the presence of papillary projections with a subjective semiquantative assessment of overall tumour vascularity, the colour score. Their model gave a sensitivity and specificity of 95.9% and 87.1%, which outperformed both the RMI (67.3% and 91.1% at a cut off of 200) and the Lerner score (96.1% and 60.7%) respectively.

In 2005, the IOTA group published two new LRMs based on data obtained from 1066 patients from nine different European centres (Timmerman, 2005). Their model included 12 variables (personal history of ovarian cancer, hormonal therapy, age, maximum tumour diameter, pain, ascites, blood flow within a papillary projection, presence of an entirely solid tumour, maximal diameter of solid component, irregular internal cyst wall, acoustic shadows and colour score). With the new model applied to a test set of 312 patients, a sensitivity of 93% and a specificity of 76% were obtained with a cut off of 0.10 for malignancy. The performance of previously published models (Jacobs' RMI, Tailor's LRM, Timmerman's LRM) on the test set was compared. None performed as well as the new LRM1 as shown in Table 1.4. Notably, they did not find CA 125 to be a statistically discriminatory variable.

Table 1.4: Comparison of the performance of different models (after Timmerman et al, 2005)

Model tested	Area under ROC	SE	Cut off	Sensitivity (%)	Specificity (%)
IOTA LRM1	0.936	0.020	0.10	92.7	74.3
IOTA LRM2	0.916	0.021	0.10	89.9	70.7
Jacobs RMI	0.870	0.028	100	78.3	79.6
Tailor LRM	0.869	0.025	0.25	63.2	88.2
Timmerman LRM	0.903	0.023	0.25	79.7	80.8

1.6.3 Artificial neural networks

Artificial neural networks (ANNs) are data analysis tools that acquire knowledge through learning. Neural networks are so called because of the similarity between their structure and that of biological brains. When a neuron fires in the physiological nervous system, the impulse is transmitted across a synapse. It is received by the dendrite of another neurone and causes an increase in the neurone's activation level. The activation level of the neurone is dependent on the incoming impulses from surrounding neurones and on the strength of the connection between them. Some connections can be excitatory, others inhibitory. Neurophysiologists have shown that animal learning is accompanied by changes in the morphology and neurotransmitter release at neural synapses. They have postulated that learning takes place by adjustment of the strengths of synaptic connections (Stein, 1982).

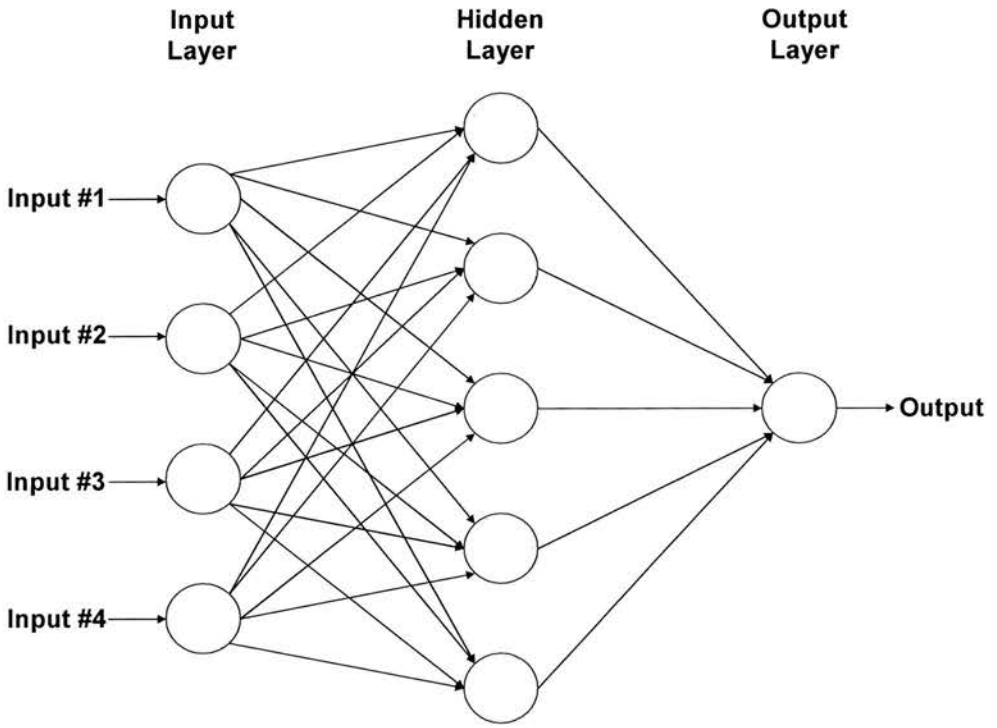
An ANN consists of at least three layers: an input layer, a number of hidden layers and an output layer (Figure 1.11). The input layer is made up of a number of nodes, each of which receives numeric input data. The data is propagated through the layers to give an output from the model. Each data parameter has its own input node. The data is processed by the node and is passed on to each of the nodes in the next or hidden layer. The magnitude of the signal passed on is dependent on the input data; the mathematical equation used to process the data, and the strength of the connection between the input node and the node of the hidden layer. The hidden layer nodes similarly transform the signal and pass it on to all the nodes in the output layer. The output layer consists of a node for each desired output.

Each connection between nodes has a weight or strength that denotes the relative importance of the nodes in the preceding layer. Initially, the weights are set at random. The model undergoes training when examples with known outcomes are presented to the network. The network can be trained using a number of algorithms. The classic method is supervised backpropagation. The difference between the expected output and the actual network output is calculated for each example as a mean square error. The size of the mean square error is backpropagated to the weights in order to minimise the error between actual and expected output. The weights between each node are adjusted and another example is presented. With each iteration, the error is recalculated and the weights readjusted. The training of the network enables optimal adjustment of the weights to give a maximum of correctly classified examples. Training can be continued either *ad infinitum* until the mean square error has been minimised, or until a set number of iterations has been performed. Minimisation of the mean square error does not always provide the best model as this can cause overfitting of

the data. This is similar to memorising the examples and gives a model that is poor at generalising with new data. Two other methods may be employed to avoid overfitting: limiting the number of hidden neurones and restricting the size of weights used by the network.

The advantages of neural networks include an ability to compute a large volume of data and to improve their performance with training. They are able to adjust to dependence between input variables by varying the weights on particular nodes. They are also able to discriminate and recognise patterns within data. However, they are unable to determine the most useful input variables and a large number of different networks must be investigated to identify the best. Another disadvantage is the "black box" nature of the hidden nodes, where the data is transformed out of sight. They can be used in a wide variety of applications including decision-making, classification and prediction.

Figure 1.11: A simple neural network



Neural networks have been used by a number of authors to predict malignancy in adnexal masses based on a set of patient data. Tailor published the first ANN designed for this purpose in January 1999. He excluded the use of more than one variable in the same model that may be co-dependent, such as age and menopausal status. After trialling 205 different ANNs, he developed a model with four demographic and sonographic variables (age, largest diameter, papillary projections and TAMXV) as inputs. The data set of 67 patients was split into a training (52) and a test set (15). A decision value of 0.45 gave an impressive sensitivity of 100% with a specificity of 98.1% (Tailor et al, 1999).

Timmerman developed two further neural networks later the same year (Timmerman et al, 1999b). The first network included the variables in

Timmerman's logistic regression model (menopausal status, CA 125, colour score and papillary proliferation). The second model included seven of a number of input variables that had been identified by logistic regression as the most relevant (menopausal status, CA 125, ascites, unilocularity, smoothness of internal wall, presence of papillary structures and bilateral masses). Around 100 networks were developed with a differing selection of inputs and network architecture. After training the networks to a minimal mean square error, the best performing model on the test set was chosen. Timmerman also produced a LRM with the same inputs based on the same training set of 116 patients and a test set of 57. The second, more complex neural network outperformed both the simpler network and the logistic regression model based on the same parameters with a sensitivity of 95.9% and a specificity of 93.5% on the complete set.

A group from Leeds developed a neural network using the inputs of age, Lerner's ultrasound score and CA 125 (Clayton et al, 1999). The ultrasound score was calculated retrospectively from a review of 144 ultrasound images by a radiologist. The exclusion criteria were not defined and there was a significant amount of missing data. They also developed a logistic regression model with the same inputs trained and tested on the same set of women. The ANN achieved a sensitivity of 95.4% and a specificity of 77.7%, outperforming the LRM on both (sensitivity 82.3%, specificity 51.2%).

Biagiotti and his group in Florence developed a number of ANNs and LRMs based on demographic, morphologic and Doppler variables. They did not use CA 125 in any of the models. They used stepwise logistic regression to identify the most discriminatory variables and included five variables in their two models (age, papillary formations, random echogenicity, PSV and

RI). They had a test population of 166 women with a cross validation set of 41. The ANN based on the five input variables gave a sensitivity of 96% with a specificity of 97.7%, significantly outperforming the LRM developed using the same variables (Biagiotti et al, 1999).

These results suggest that an ANN performs better with raw morphological data rather than data combined into an ultrasound score such as used in Clayton's model. Clayton's model may have underperformed due to retrospective collection of ultrasound data and the large amount of missing data. The ANNs appear to have outperformed the published LRMs but as yet have not been satisfactorily validated.

1.6.4 Decision trees

Decision trees or classification and regression trees (CART) are a method of dividing cases into classes using one or more predictor variables. Their aim is to predict the membership of a case based on previous examples. CART are hierarchical and recursive in nature. They are able to split data a number of times using the same variable as opposed to a linear analysis which allows only a single split which could leave a substantial amount of the information in the predictor variables unused. They are easily displayed graphically and are easy to interpret. They are used in a wide ranging number of applications including classification in botany, psychology (decision theory) and data structuring in computer science.

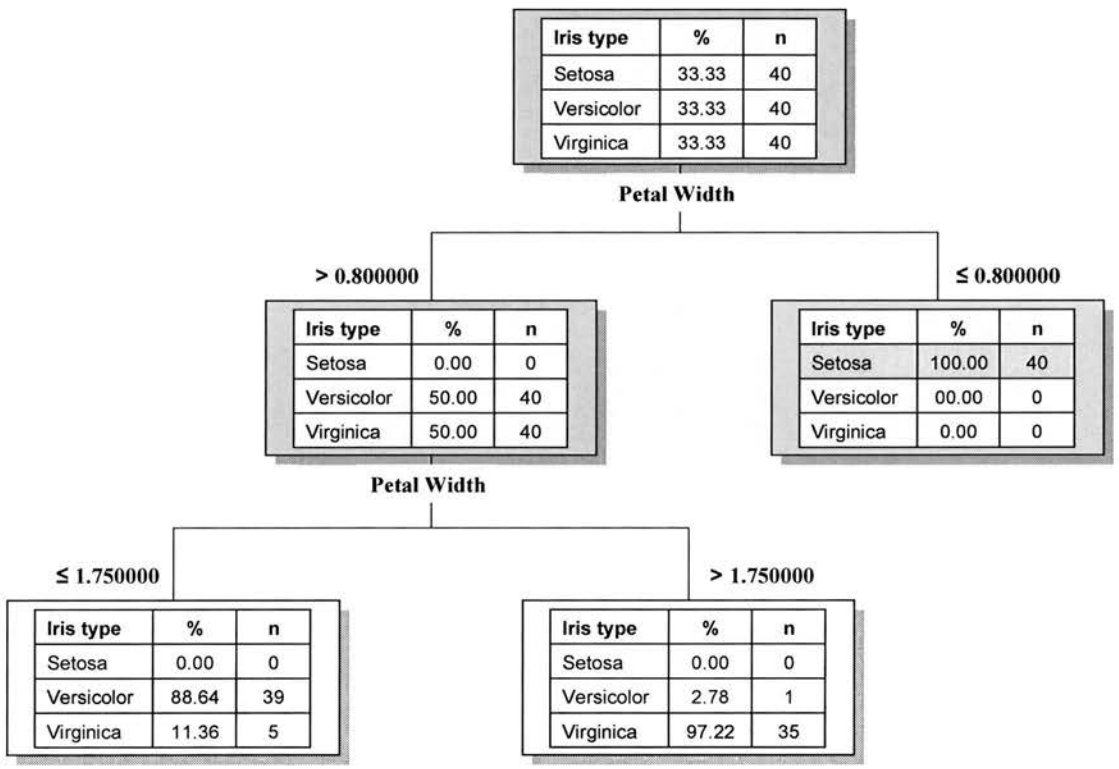
Fisher described a classification example to place three different types of iris flowers (Setosa, Versicol, and Virginic) into their classes (Fisher, 1936).

These vary in breadth of petals: Setosa have petals of breadth less than 8mm; Versicol between 8 and 17.5mm and Virginic greater than 17.5mm. Linear analysis can divide two of the classes but not the third. A decision tree can be developed to discriminate between the classes with two logical rules: if the petal is narrower than 8mm, it is Setosa, if broader then Virginic or Versicol. If the petal is then narrower than 17.5mm, it is Versicol, if broader it is Virginic (Figure 1.12). Breiman described the classic computational algorithms for classification trees in 1984 (Breiman et al, 1984).

A decision tree starts with a single parent node containing all the cases. The data is then divided by a rule into two groups that pass to two child nodes. These nodes then divide their cases based on further rules until the cases are divided into their constituent groups at the final or terminal nodes. Decision trees are so called because of their similarity in structure to a tree with a single trunk branching many times until the leaves or terminal nodes are reached. The trees are able to examine the effect of each variable, one at a time. They are also flexible in their ability to deal with both categorical and ordinal data.

They have been successfully applied to the diagnosis of ovarian malignancy using tumour markers. A sensitivity of 90.6% and a specificity of 93.2% were obtained with the use of 5 markers which was significantly better than the use of CA125 alone. A number of paradoxical splits were identified in the tree, where a small number of benign masses with a high level of one marker could only be distingished from malignant masses by virtue of having a higher serum level. Pruning these nodes reduced the sensitivity to 82.3% (Woolas et al, 1995).

Figure 1.12: A decision tree for the classification of irises. There are five nodes with the three terminal nodes outlined in pink. The data is divided by two logical rules of petal width.



1.6.5 Validation of prognostic models

Validation of a model is defined as establishing that it works satisfactorily for patients other than those from whose data it was derived (Altman and Royston, 2000). A model can be invalid for two main reasons: statistical invalidity and clinical invalidity. Statistical invalidity is when the model is not the best that can be found, given the prognostic information. Clinical invalidity arises when the prognostic information is too weak to permit

development of a model that can accurately divide the data set into clinically useful groups.

Validation can thus be statistical or clinical for any given model. A clinically validated model is defined as one that performs satisfactorily on a new data set according to context-dependent statistical criteria laid down for it (Altman and Royston, 2000).

There are a number of reasons why a model may be invalid. Fitting the model too closely to the training set of variables on which the model is developed may produce an overoptimistic model. If the model is then tested on the same data set then the performance is likely to be overestimated. This risk can be minimised by reducing the number of variables included in the model, either using clinical experience or statistical methods such as logistic regression modelling.

Bias may be introduced when developing the model. This can be due to missing data, which may not be missing randomly, or poorly defined inclusion and exclusion criteria. A small sample size for model development gives a low signal-to-noise ratio. This can lead to important variables being missed from the model and unimportant ones being included. A number of authors have suggested that the number of variables included in a model should be relative to the size of the data set. Harrell concluded that for regression modelling the size of the data set should be at least ten times the number of potential prognostic variables that could be included in the model (Harrell et al, 1984). Peduzzi reported simulation studies that showed that parameter estimates in logistic regression models are unreliable when this ratio is less than 10 (Peduzzi et al, 1996).

For a model to be useful in clinical practice, it must be transportable to other locations. Dissimilarity between patients in different centres may lead to poor model performance. If a model includes all the important variables and has attached sufficient weight to each then it should perform equally well in centres with a different case mix. However, it is not possible to confirm that a model indeed does contain all the important variables. It is important to validate a model to ensure its performance is reproducible.

Validation of a model may be internal, temporal or external (Altman and Royston, 2000). Internal validation involves splitting the dataset to test the model on selected cases. This may be by division into a training set for model development and test set to assess performance or by cross validation by a leave-one-out method. Temporal validation is the collection of a subsequent dataset in the same centre to test the model on. This is a prospective method but is otherwise the same as splitting the original dataset into two cohorts divided by time. Neither internal nor temporal validation assesses how well a model will perform in a different population. This can only be answered by external validation through the application of the model to a dataset from a different centre.

1.6.6 Prospective validation of models

Davies et al performed a retrospective temporal validation of Jacobs' RMI in 1993. The population consisted of 87 benign masses and 37 malignancies, 28 of which were primary epithelial ovarian cancer. Using the same cut off of 200, the RMI gave a sensitivity of 87%, a specificity of 89% and a positive predictive value of 75% (Davies et al, 1993).

Morgante et al retrospectively evaluated the two risk of malignancy indices in a new population of 124 Italian women from Siena. This external validation demonstrated that Tingulstad's model significantly outperformed CA 125, menopausal status and Jacobs' RMI. They calculated the performance of the models at different cut off values (Table 1.5). The optimal cut off value for Tingulstad's RMI was 125 in the Italian population as opposed to the originally described level of 200 (Morgante et al, 1995).

Table 1.5: Performance of Jacobs' and Tingulstad's RMIs at different cut off levels (after Morgante, 1999)

	Sensitivity (%)		Specificity (%)	
Cut off	**Jacobs' RMI**	**Tingulstad's RMI**	**Jacobs' RMI**	**Tingulstad's RMI**
25	97	97	60	57
50	94	94	75	70
80	81	90	80	78
100	77	84	90	86
125	74	81	92	90
150	65	77	94	91
200	58	74	95	93
250	54	65	96	94

Aslam et al carried out a prospective external validation of both Jacobs' and Tingulstad's RMIs (Aslam et al, 2000b). She compared this to the performance of Tailor's model, which was developed at King's College Hospital, London. She found that all three models performed less well than

in the original publication as expected. However, Tailor's model performed particularly poorly with a sensitivity of 43% and a specificity of 92%. This result was surprising as the validation was temporal rather than external although the ultrasonographers were different from the original publication. Aslam suggested that this was due to a difference in case mix between the two studies. The malignancies in Aslam's dataset included 22% non-epithelial cancers, a type that was absent in Tailor's dataset. His model placed a high weighting on the presence of papillary projections, a feature that is often absent in non-epithelial and some borderline tumours. Of the other two models, their performance was similar with a sensitivity of 74% and specificities of 92% (Jacobs) and 89% (Tingulstad). (Table 1.6)

Table 1.6: Prospective performance of Jacobs' RMI

Group	Year of Publication	Validation	Cut-off Value	Sensitivity (%)	Specificity (%)
Davies	1993	Temporal	200	89	87
Tingulstad	1996	External	200	71	96
Morgante	1999	External	200	58	95
Aslam	2000	External	200	74	92
Mol	2001	External	200	90	61
Timmerman	2005	External	100	78.3	79.6

Three logistic regression models were tested prospectively. Alcazar and Timmerman's models were applied to a population from King's College and

compared to the performance of Tailor's model. Timmerman's model performed best with an area under the curve of 0.86. The sensitivities of the other two models were disappointing as shown in Table 1.7. The three models were combined but this gave no improvement on Timmerman's results (Aslam et al, 2000c).

Table 1.7: Prospective performance of logistic regression models (after Aslam, 2000c)

Model	Sensitivity	Specificity	Area under ROC curve	95% Confidence Interval
Tailor	45	93	0.86	0.77-0.94
Alcazar	9	99	0.69	0.58-0.81
Timmerman	73	91	0.85	0.89-0.98

Two logistic regression models were compared to the performance of pattern recognition in a prospective study (Valentin et al, 2001). Both Tailor's and Timmerman's models performed similarly (AUC 0.87 and 0.84 respectively) but were less accurate than pattern recognition.

A paper describing the prospective external validation of Timmerman's model was published in 2001(Mol et al, 2001). The experimental design of this study was queried by the original group (Timmerman et al, 2001). The data was collected between 1991 and 1998; however the end points for analysis were not described until 1999. The colour score is a subjective measurement of tumour vascularity and there was a dramatic improvement in sensitivity of ultrasound equipment for the detection of colour flow between 1990 and 1999. In addition, the CA 125 II immunoradiometric assay

used in the original publication was only available from 1994. The definition of both papillary projections and menopausal status varies between different models. It is unclear whether Mol et al retrospectively examined and reclassified their data prior to the insertion in each model. Timmerman et al have suggested that their model should be evaluated by independent external assessors in well-designed, prospective studies.

1.6.7 Small masses

Small, asymptomatic adnexal masses are commonly detected in women undergoing a scan for other indications. These indications include ovarian cancer screening and irregular bleeding both in and outside pregnancy. These masses can cause concern for the patient but may also lead to further investigations and, in many cases, unnecessary surgery. The potential for doing harm to this group of asymptomatic women is significant. In Campbell's ovarian screening study in 1989, 65 masses were detected in the healthy screening population for each cancer (Table 1.8). These women underwent a staging laparotomy and its considerable attendant morbidity for no personal benefit. There is little literature to guide management in this increasingly common group of masses.

Bailey et al examined 506 cysts with a diameter smaller than 10 cm in asymptomatic women over the age of 50 undergoing ovarian cancer screening. Half were unilocular; the remainder had solid cystic tumours. Half of the cysts resolved spontaneously. Of the women who underwent surgery, none of the unilocular cysts was malignant as opposed to 8% of the complex cysts.

Table 1.8: Prevalence of adnexal masses and primary ovarian cancer in asymptomatic women screened using ultrasound (after Valentin, 1999b)

Study	n	Abnormal ultrasound finding		Operated on due to abnormal ultrasound		Primary ovarian cancer	Operations per cancer
		n	%	n	%		
Campbell 1989	5479	326	5.9	326	5.9	5	65
van Nagell 1990	1000	31	3.1	24	2.4	0	∞
van Nagell 1991	1300	33	2.5	27	2.1	2	14
DePriest 1993	3220	44	1.4	44	1.4	3	15
Schulman 1994	2117	202	9.5	18	0.9	1	18
Kurjak 1994	5013	424	8.5	38	0.8	4	10
Parkes 1995	2953	87	2.9	9	0.3	1	9
All studies	**21082**	**1147**	**5.4**	**486**	**2.3**	**16**	**30**

Ferrazzi developed a morphological scoring system for masses with a mean diameter of less than 5 centimetres in 1997 (Table 1.9). This was adapted from Sassone's scoring system to decrease the false positive results found with dermoids and haemorrhagic cysts.

Table 1.9: Scoring system for small masses (after Ferrazzi 2005)

Score	Capsule	Septa	Papillary Excrescences	Echogenicity
1	<3mm	Absent	Absent	Anechoic
2	>3mm	Thin (<3mm)		Low echogenicity / ground glass
3		Thick (>3mm)		
4	Irregular, solid		<3mm	With solid areas
5	Irregular, not applicable		>3mm	Inhomogeneous, solid

He compared it to the performance of four other scoring systems on small masses (Table 1.10). The same group then tested it prospectively in a multicentre Italian study with 677 cases. With a score over eight denoting malignancy, they found a sensitivity of 92% with a specificity of 76.9% (Ferrazzi, 2005). There is no published literature on the efficacy of logistic regression models or neural networks on this group of small, clinically undetected masses.

Table 1.10: Comparison of the performance of Ferrazzi's scoring system to other morphological scores on small masses (after Ferrazzi, 2005)

	Ferrazzi	Granberg	Sassone	De Priest	Lerner
Cut off value for malignancy	>8	>1	>8	>4	>3
Diagnostic accuracy (%)	69	53	63	52	63
p value (difference between Ferrazzi's model)	-	0.0002	0.349	0.0001	0.142

1.7 Hypotheses

This thesis will investigate the following hypotheses:

1. Published models can be effectively applied to small asymptomatic masses

2. Subjective assessment can be used to diagnose ovarian cancer in small asymptomatic masses.

3. New tumour markers will improve the preoperative diagnosis of ovarian cancer

4. Prospective testing of published models will confirm their accuracy in a new population.

1.8 Aims of the Thesis

1. To investigate the use of new tumour markers in the preoperative diagnosis of ovarian cancer. To develop new models incorporating tumour markers to diagnose ovarian cancer. This is addressed in chapter 3.

2. To externally validate published models in a new study population. To compare the performance of these models to subjective assessment and to the models developed in this thesis. This is addressed in chapter 4.

3. To investigate the differences between small asymptomatic masses and large masses. To investigate the accuracy of published models on the diagnosis of malignancy in small masses. This is addressed in chapter 5.

Chapter 2

General Methods

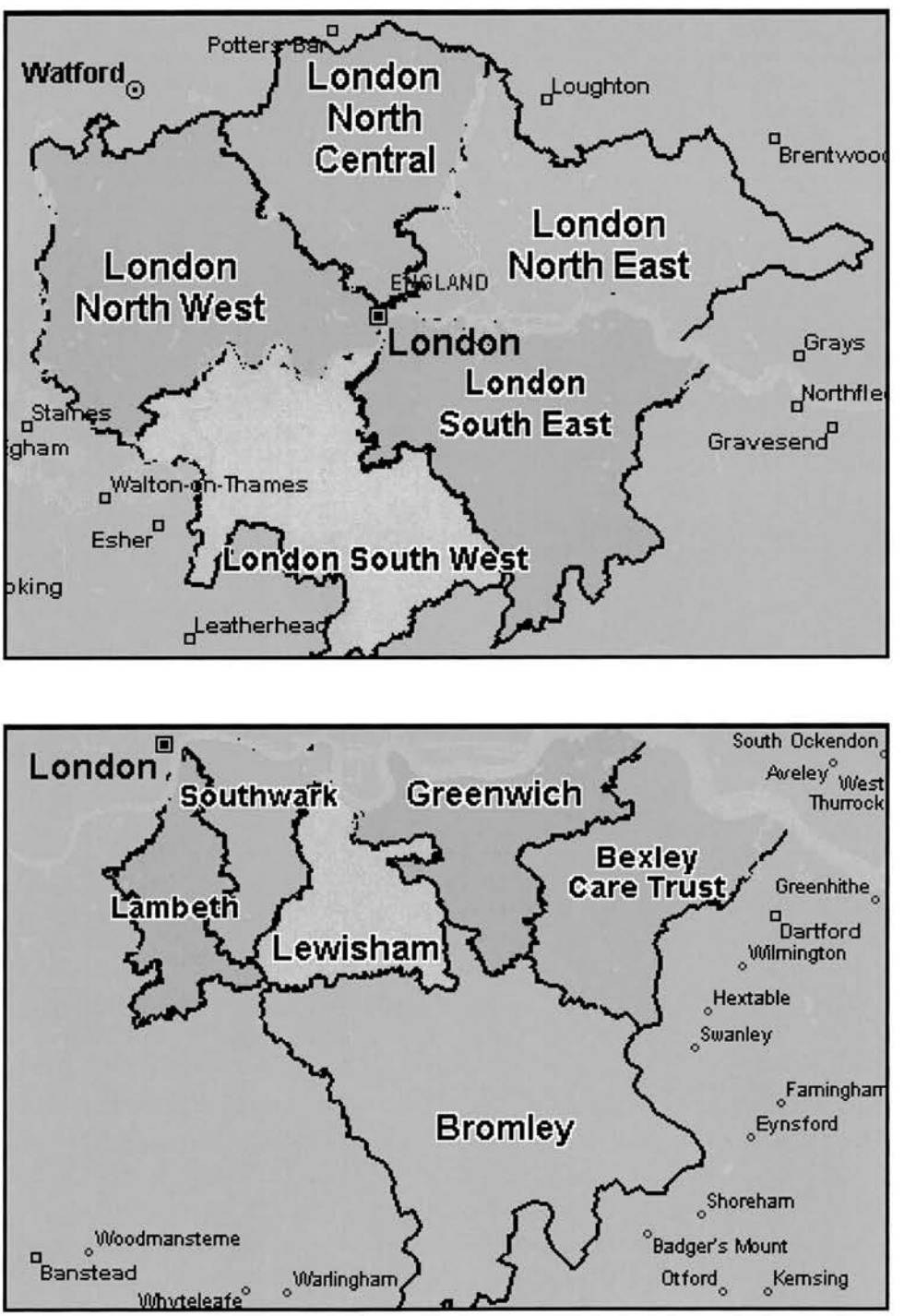
2.1 Setting, The Early Pregnancy & Gynaecology Assessment Unit (EPAGU), King's College Hospital, London

The work contained in this thesis was carried out in the Early Pregnancy and Gynaecology Assessment Unit (EPAGU) of King's College Hospital between August 1999 and August 2001. King's College Hospital is a large London teaching hospital situated in Camberwell, which primarily serves the Lambeth, Southwark and Lewisham Health Authorities and their constituent Primary Care Group Trusts (Figure 2.1). Whilst the bulk of the clinical workload is provided by referrals from the local community, King's also acts as a tertiary referral centre, having both a national and international referral pattern. The local population of 700 000 is one of the most deprived in the United Kingdom and has a large immigrant population of West Africans, East Africans and Afro-Caribbeans. More recently there has also been an increase in the number of refugees from Eastern Europe.

The hospital itself has 900 beds and employs over 4000 staff with an annual budget of £264 million. In 2001, over 450,000 patients were treated. The Women's and Children's care group consists of the Department of Obstetrics and Gynaecology as well as the Department of Paediatrics. Within the gynaecology unit, there are specialist services in acute gynaecological scanning, oncology, colposcopy, menopause, urogynaecology, assisted conception and family planning. In 2001, a total of 9386 women were referred to the EPAGU, 5024 women with problems of early pregnancy and 4452 non-pregnant women with gynaecological complaints. The EPAGU offers a tertiary referral ultrasound service for southeast London. At the time

of this thesis, the regional cancer centre was located at King's College Hospital and the EPAGU provided its diagnostic ultrasound service.

Figure 2.1: Local Health Authorities in London and South East London



2.2 Subjects

All women were invited to participate following referral to the Early Pregnancy and Gynaecology Assessment Unit. Referral was via their General Practitioner, Accident and Emergency or a hospital consultant to investigate a suspected pelvic mass. A woman was included if she was found to have a non-physiological pelvic mass that was found to originate from the adnexa as judged by ultrasonography. In the case of bilateral tumours, the mass with the most suspicious morphological features as judged by the ultrasonographer was included. In the case of a similar appearance of both masses, the larger was documented. Informed consent was obtained prior to ultrasound examination and venepuncture.

Women were excluded from the study for the following reasons: pregnancy, refusal of transvaginal ultrasonography, lack of fitness for or refusal of surgery, previous bilateral oophorectomy or the finding of a benign mass in one ovary and a borderline or invasive mass in the other.

The women included in Study 2 are the same population as those in Study 3. A number of these women did not consent to venepuncture so were not included in Study 1, leading to a smaller number of women in the tumour marker arm of the study.

Approximately a third of the women in the study were referred to the Early Pregnancy and Gynaecology Assessment Unit for a tertiary level scan. Referrals were taken from surrounding units including St Thomas' Hospital,

Lewisham, Greenwich, Epsom and St Helier's. Due to the difficulty in obtaining outcome details from other units, details of the patient pathway, MRI and CT results and specific operative findings were not recorded.

2.3 Data collection

A database file was set up using Microsoft Excel for Windows (Redmond, WA, USA) to facilitate data entry and retrieval. Age, parity and the menopausal status of each woman were recorded. A medical history was taken including a personal history of breast or ovarian cancer, hysterectomy, hormonal use and the presence of pain or irregular vaginal bleeding. Women over the age of 49 who had undergone a hysterectomy were defined as postmenopausal. Pain was graded on a scale of 0 to 3 with 0 being no pain, 1, mild pain, 2, moderate pain necessitating analgaesia and 3, pain severe enough to limit normal activities. A family history of relatives affected by breast or ovarian cancer was documented.

2.4 Ultrasound examination

All women were scanned by the same ultrasonographer (ACL). All ultrasound examinations were performed in the dorsal lithotomy position with an empty bladder. A 5MHz end-firing curved-array transvaginal probe with B-mode and Doppler facilities was used, incorporating a field of view of 90^0 (Aloka SSD-5000, Aloka Co. Ltd, Tokyo, Japan). The wall filter was set at 50 Hz and the pulsed Doppler sample volume size was set at 1.0 mm. The probe was cleaned with a hard surface disinfectant between patients and covered with a probe cover throughout each examination. All ultrasound examinations were performed within one month of surgery.

A detailed gray-scale examination of the pelvis was performed with inspection of the uterus and contralateral ovary. The origin of the adnexal mass was determined if possible and classified as ovarian, tubal, other or uncertain. If the origin was doubtful, a sliding organs technique was employed. Gentle pressure was applied with the transvaginal probe with the other hand used to palpate the abdomen. This allows separate structures to slide apart and identification of the origin of a mass.

The adnexal mass was judged to be the part of the mass or ovary that was inconsistent with normal physiology. In the case of a persistent cyst within an otherwise healthy ovary, the cyst was regarded as the mass and the cyst and the ovary were measured separately. If the mass was identifiably distinct from the ovary (e.g. a fimbrial cyst) then both the ovary and the mass were again measured separately. If there was no recognisable ovarian tissue,

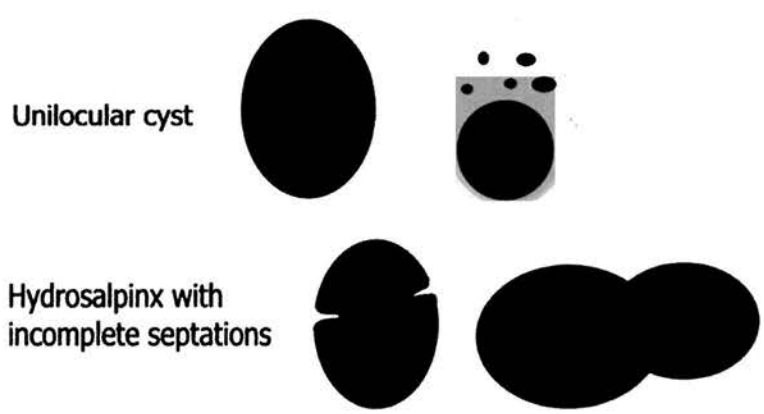
then the ovary and the mass were indistinguishable and so the measurements of both ovary and mass were the same. Measurement of the volume of the mass (excluding healthy ovarian tissue) and the ovary was made in three perpendicular planes using the formula for a prolate ellipsoid.

$$\pi/6 \times D_1 \times D_2 \times D_3$$

2.4.1 Morphology

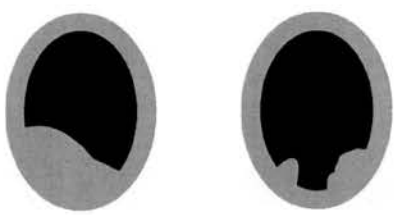
The mass was classified into one of six groups according to the International Ovarian Tumour Analysis (IOTA) definitions (Timmerman et al 2000): unilocular, unilocular-solid, multilocular, multilocular-solid, solid or unclassifiable. A unilocular mass was one containing a single cyst, which did not contain a solid component (Figure 2.2).

Figure 2.2: Examples of unilocular cysts



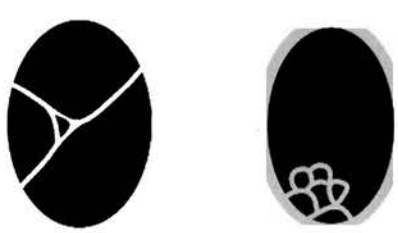
A unilocular solid mass was a unilocular cyst with a measurable solid component or at least one papillary projection (Figure 2.3). This category may include a hydrosalpinx with a "beads on a string" appearance if the solid parts are greater than 3mm in height.

Figure 2.3: Examples of unilocular solid cysts



A mass was defined as multilocular if it contained at least one complete septum, which divided the cyst into two or more locules and had no measurable solid component (Figure 2.4).

Figure 2.4: Examples of multilocular cysts



A multilocular cyst with a measurable solid component or at least one papillary projection was defined as multilocular solid (Figure 2.5).

Figure 2.5: Examples of multilocular solid cysts

If more than 80% of the mass was solid, it was classified as solid. A solid mass could therefore contain small cystic structures within it. To distinguish between a solid mass and a haemorrhagic cyst, two methods were employed. Firstly, the mass was gently moved with the probe and the mass contents were inspected visually. If the mass were composed of clot then the 'jelly-like' clot would be seen to move after pressure from the transducer. If it were solid then no movement would be seen. The second method was to survey the structure with colour Doppler. The presence of flow within vessels was diagnostic of solid tissue. The absence of flow was not informative and in the case of doubt, the mass was recorded as solid. A solid mass could contain papillary projections projecting into the cyst cavities (Figure 2.6).

Figure 2.6: Examples of solid masses

Some masses could not be classified into one of the other three categories and therefore were recorded as unclassifiable. An example of this is a dermoid cyst with a "tip of the iceberg sign" demonstrating acoustic shadowing (the loss of an acoustic echo behind a sound reflecting surface) causing the structures behind it to be obscured.

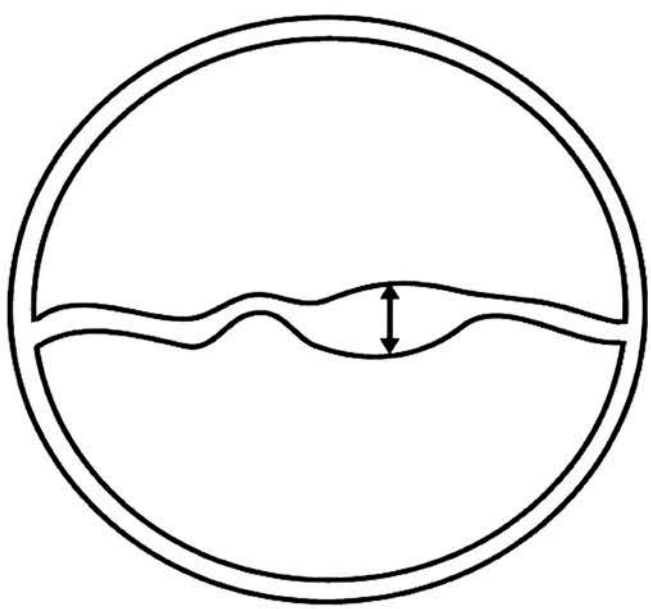
2.4.2 Echogenicity

The echogenicity of the cyst contents was described as anechoic, low level (homogeneous, low-level echogenicity as seen in mucin), ground glass (homogeneous, echogenic fluid as typified by endometriotic cysts), haemorrhagic (the presence of a stellate or 'jelly-like' clot with visible fibrin strands) or of mixed echogenicity (as often seen in a mature teratoma). If a solid tumour contained a cystic area, then its contents were described. The internal cyst wall was described as smooth or irregular. If the mass was solid then its external contour was described as smooth or irregular.

2.4.3 Septum

A septum was defined as a strand of tissue dividing the cyst cavity and its thickness measured perpendicular to the cyst wall at its widest point save at its interface with the cyst wall (Figure 2.7). An incomplete septum (as seen in a hydrosalpinx) was defined as a strand of tissue dividing the cyst cavity that was incomplete in some scanning planes. If incomplete septa only were present, the cyst was defined as unilocular despite its multilocular appearance in some sections (Figure 2.2).

Figure 2.7: Measurement of septum thickness



2.4.4 Papillary projection

A papillary projection was defined as a solid projection from the cyst wall into its cavity greater than or equal to 3mm in height (Figure 2.8). The number of separate papillary projections was recorded and their size measured in three planes (Figure 2.9). The projections were described as either smooth or irregular in outer contour. The presence of detectable flow within the projections was recorded. If there was uncertainty as to whether a solid projection from the cyst wall was a papillary projection or an incomplete septation (e.g. the cogwheel excrescences seen in hydrosalpinges), then the projection was classed as a papillary projection if its height exceeded 3mm. The deposits of dense fluid at the base of endometriotic cysts were not included as papillary projections but the internal walls were classified as irregular.

Figure 2.8: A serous cystadenoma consisting of a unilocular anechoic cyst with a single papillary projection (arrowed).

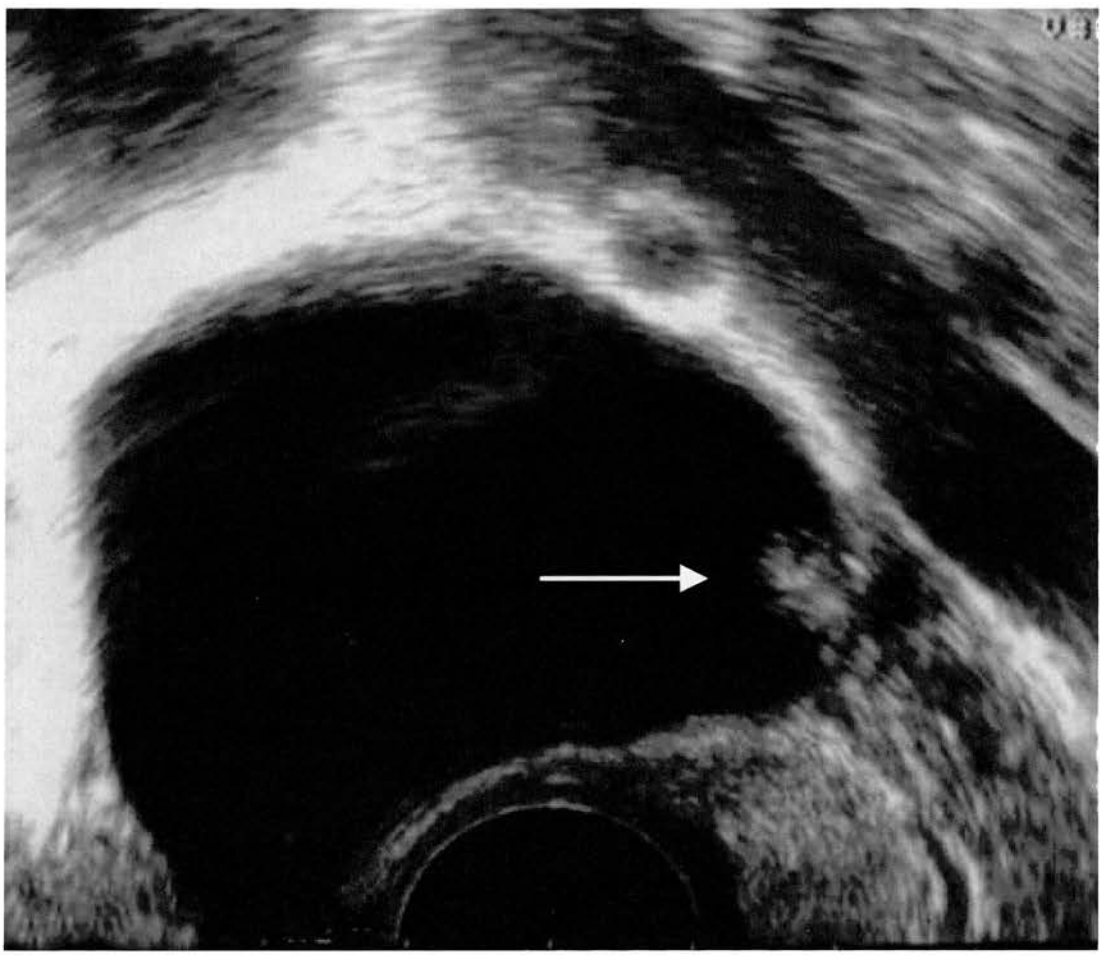
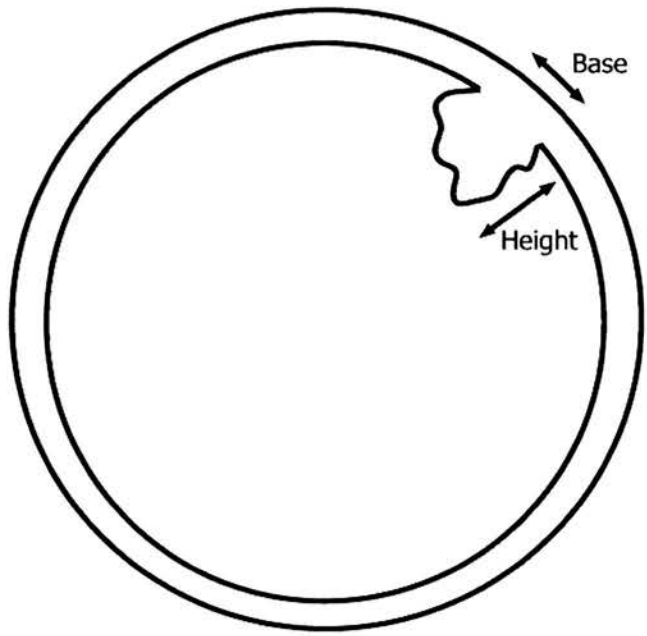
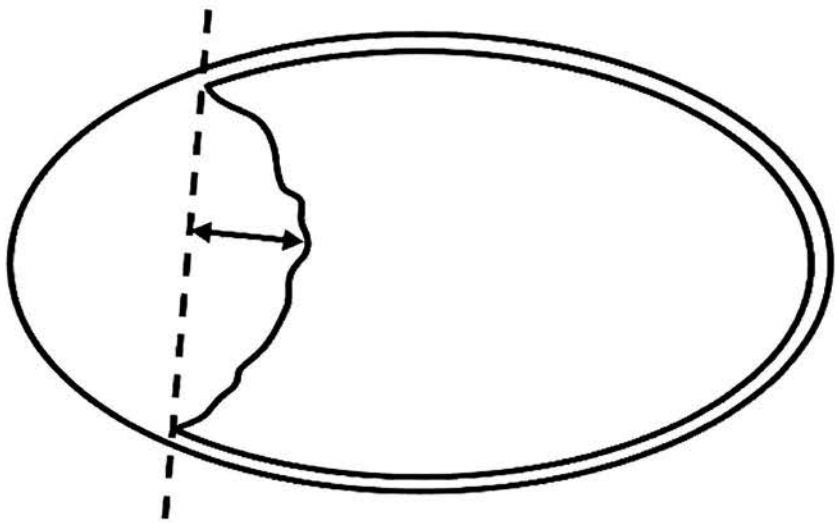


Figure 2.9: Measurement of a papillary projection



In some cases, it was difficult to measure a projection arising from a solid area within a mass. An imaginary line was used to denote the base of the projection from which its height could then be measured (Figure 2.10).

Figure 2.10: Measurement of the height of a papillary projection arising from a solid area.



The presence of solid components was noted and the largest solid area was measured in three dimensions.

2.4.5 Acoustic shadowing

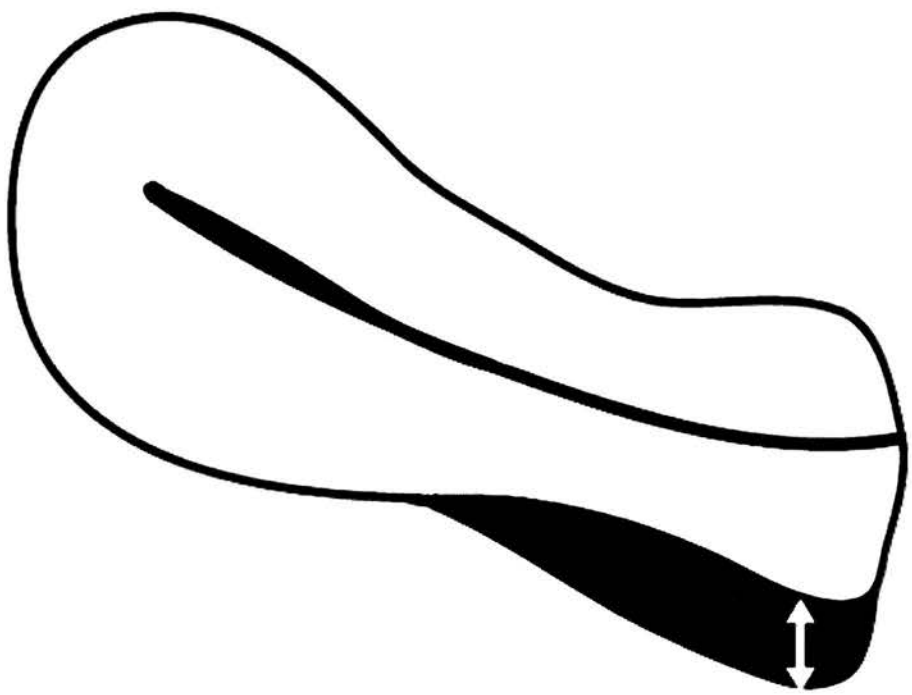
The presence of acoustic shadowing defined as the loss of an acoustic echo behind a sound reflecting surface (e.g. the fluid/fat interface of a dermoid cyst) was recorded.

2.4.6 Ascites

Ascites was determined as free fluid outside the pouch of Douglas. The depth of free fluid within the pouch of Douglas was measured

anteroposteriorly in the sagittal plane (Figure 2.11). The fluid was judged to be free rather than encapsulated (e.g. peritoneal pseudocyst) if it could be moved out of the pouch of Douglas by gentle pressure with the probe.

Figure 2.11: Measurement of free fluid in the pouch of Douglas



2.4.7 Colour Doppler Imaging

Colour Doppler imaging was chosen to investigate the vascularity of the mass. This method was validated in a population of 56 women with corpora lutea and 69 women with known adnexal masses. The reproducibility and diagnostic accuracy of colour Doppler imaging (CDI) and colour Doppler energy (CDE) were compared. There was no significant difference in the diagnostic performance of the two modalities. The reproducibility of CDI

was superior to that of CDE for all four Doppler variables (time-averaged maximum velocity (TAMXV), peak systolic velocity (PSV), resistance index (RI) and pulsatility index (PI)) as shown in Table 2.1 (Tailor 1998).

Table 2.1: Reproducibility of blood flow indices with CDI and CDE (after Tailor, 1998)

	Coefficient of Variability		Intraclass Correlation Coefficient	
Measurement	**CDI**	**CDE**	**CDI**	**CDE**
PSV	14.0	22.8	0.984	0.977
TAMXV	13.4	21.7	0.986	0.978
PI	14.8	17.6	0.888	0.885
RI	9.7	10.6	0.889	0.872

The whole mass was surveyed using CDI. The power, gain and pulse repetition frequency were set for the maximum sensitivity for detection of low velocity flow. Low velocity signals were filtered out by slowly increasing the pulse repetition frequency and flow analysis was concentrated on the highest velocity signals. These high velocity signals were also identified by the phenomenon of aliasing. The pulsed Doppler range gate was placed across the highest velocity signal and the Doppler flow waveform recorded. Minute adjustments to the angle of the probe were made until the audible signal was optimal. This was considered to be the optimal angle of insonation for the vessel at that particular location and no angle correction

was made. The Doppler parameters recorded were the time-averaged maximum velocity (TAMXV), peak systolic velocity (PSV), resistance index (RI) and pulsatility index (PI). These parameters were calculated electronically from a smooth curve fitted to the Doppler waveform. If multiple areas of high velocity were identified, the set of results with the highest TAMXV was used.

A subjective estimation of the mass's vascularity was made on a scale ranging from 1 to 4. This colour score was described by Timmerman in 1999 and has been used by a number of ultrasonographers (Valentin, 2002, Timmerman 1999 and the IOTA group). No blood flow within the solid areas of the mass was denoted as 1, 2 was given for minimal flow, 3 when moderate flow was present and 4 when the mass appeared highly vascular. This score was given once only for the mass as a whole.

At the end of the scan, the operator made a subjective decision as to whether the mass was malignant or benign. It was then categorised as benign, borderline or invasive. The likely histological type of the mass based on a subjective impression during the scan was also noted e.g. endometrioma, serous cystadenoma, mucinous cystadenocarcinoma.

Multiple photographic prints of the uterus, adnexal mass and contralateral ovary were taken.

2.5 Venous blood samples

Blood samples were taken with informed consent from a peripheral vein. The blood was centrifuged within 2 hours for 15 minutes at 300 rpm and aliquots of the supernatant stored at –20^{0}C until analysis. Each aliquot was thawed only once with any remaining serum discarded after thawing.

2.6 Measurement of tumour markers

All assays were performed in duplicate and any samples where either of the duplicate values did not fall into within 10% of the other were reassayed. When a quality control value differed by greater than 10% from previous assay means, the assay was repeated. The details of the assays are documented in Appendix V.

2.7 Histology

Surgery was carried out either laparoscopically or at laparotomy depending on the surgeon's judgement. The whole tumour was removed in the majority of cases, however in the case of disseminated ovarian malignancy or endometriosis, representative biopsies were taken.

Tumours were staged by the attending gynaecologist according to the International Federation of Gynaecology and Obstetrics (FIGO) classification (Appendix I). Histological specimens were mounted on blocks and one section per centimetre was taken for examination by a pathologist. The tumours were classified in accordance with the World Health Organisation classification and were graded as well-differentiated (Grade 1), moderately differentiated (Grade 2) or poorly differentiated (Grade 3).

As previously noted, all women underwent a repeat scan prior to surgery. At this time, eleven masses were noted to have resolved spontaneously. Surgery was cancelled for these women and the outcome of spontaneous resolution was recorded.

2.8 Statistical analysis

The Statistical Package for Social Sciences version 10.0 (Statistical Analysis Systems, Chicago, Illinois) was used to analyse data. Details of individual tests are documented in the relevant chapters. The level of statistical significance was chosen as $p<0.05$.

2.9 Prospective testing of published models

Six models were tested prospectively on the data set of 170 women. Four of the models have been tested prospectively by other authors: Jacobs' Risk of Malignancy Index, Tingulstad's Risk of Malignancy Index, Tailor's logistic regression model and Timmerman's logistic regression model (Aslam 2000, Valentin 2001, Mol 2001). Although, Timmerman's neural network model has been tested prospectively by Mol in 2001, significant concerns over the study methodology have been raised by the group who published the original model (Timmerman at al, 2001). The logistic regression model published by the IOTA group in 2006 has not been prospectively validated. The details of all six models are shown in Table 2.4. In each model, 0 denotes the absence of a variable and 1, its presence. The symbol e is the mathematical constant and base value of natural logarithms.

Statistical analysis and the construction of ROC curves were carried out using SPSS for Windows as previously described.

Table 2.2: Prospectively tested models: a summary of their equations, variables and year of publication

Author	Publication Year	Variables Included in Model	Equation
Jacobs	1990	Ultrasound score (0 to 3) Menopausal status (1 or 3) CA 125 (kU/L)	U x M x CA 125
Tingulstad	1996	Ultrasound score (0,1 or 4) Menopausal status (1 or 4) CA 125 (kU/L)	U x M x CA 125
Tailor LRM	1997	Age (years) TAMXV Papillary projection score (0 or 1)	$p = 1/(1 + e^{-z})$ where z = (0.1273 x age) + (o.2794 x TAMXV) + 4.4136 x papillary score – 14.2046
Timmerman LRM	1999	Menopausal status (0 or 1) CA 125 (kU/L) Ascites (0 or 1) Unilocularity (0 or 1) Smooth internal wall (0 or 1) Papillary projection score (0 or 1) Bilateral masses (0 or 1)	$p = 1/(1 + e^{-z})$ where z = (0.5948 x menopausal) + (0.0205 x CA 125) + (0.5446 x ascites) – (0.762 x unilocularity) – (1.1606 x smooth) + (1.5409 x papillary) + (0.7633 x bilateral) – 1.0889

Author	Publication Year	Variables Included in Model	Equation
Timmerman NN	1999	Menopausal status (0 or 1) CA 125 (kU/L) Ascites (0 or 1) Unilocularity (0 or 1) Smooth internal wall (0 or 1) Papillary projection score (0 or 1) Bilateral masses (0 or 1)	$p = 1/(1 + \exp(z_y))$ where $z_y = 2.9753\ h_1 + 4.1980\ h_2 - 3.8616$ in which $h_1 = 1/(1+ \exp(z_1))$ with z_1 = -1.0792menopausal + 1.9383 CA 125 + 0.71242 ascites – 1.2664unilocular – 1.3741smooth + 0.8298papillary + 1.5316bilateral – 0.5485 $h_2 = 1/(1+ \exp(z_2))$ with z_2 = 1.0766menopausal + 0.1376CA 125 + 1.0112ascites – 0.8320unilocular – 1.6941smooth + 2.9541papillary + 1.4654bilateral – 1.8129
IOTA LRM	2006	Personal history of ovarian cancer (0 or 1) Current hormonal therapy (0 or 1) Age (years) Maximum mass diameter (mm) Pain during exam (0 or 1) Ascites (0 or 1) Blood flow within papillary projection (0 or 1) Purely solid tumour (0 or 1) Maximal diameter of solid component (<50mm) Irregular internal cyst wall (0 or 1) Acoustic shadows (0 or 1) Colour score (1 to 4)	$p = 1/(1 + e^{-z})$ where z = -6.7468 + 1.5985 x history of ovarian cancer – 0.9983 x hormonal therapy + 0.0326 x age + 0.00841 x maximum mass diameter – 0.8577 x pain during exam + 1.5513 x ascites + 1.1737 x blood flow within papillary projection + 0.9281 x purely solid tumour + 0.0496 x maximal diameter of solid component + 1.1421 x irregular internal cyst wall – 2.3550 x acoustic shadow + 0.4916 x colour score

2.10 Decision tree analysis

Data was analysed using a Classification and Regression Tree (CART) software package Statistical Package for Social Sciences Answer Tree (Statistical Analysis Systems, Chicago, Illinois). Two types of trees were 'grown', one with three coded outputs or terminal nodes: benign, borderline and invasive, the second with two terminal nodes: benign and malignant. The stopping rules for the iterative process were that the tree should have a maximum of five levels, a minimum of 10 cases were to be present for a split to be calculated and any given split should not generate a group with less than 5 cases. Six variables were used for construction of the decision tree and these included age, CA 125, CA 19-9, CA 72-4, CA 15-3 and HER-2/neu. All were entered as continuous variables. The accuracy of the models was judged by the risk estimate, which gives the proportion of cases classified incorrectly.

V fold cross validation was used to identify the trees with the best average accuracy at predicting values. This technique for assessing the trees allows a random sample of the data to be tested on the tree and the generated predictions compared to the known outcomes. This is done multiple times to identify the smallest effective tree. 10 folds of cross-validation were used.

2.11 Artificial neural network analysis

The dataset was split randomly into a training set comprising approximately two thirds of the cases and a test set containing the remainder. The calculations during the development of the neural networks were programmed in Thinks Pro 1.05, (Logical designs Consulting Inc). All calculations of receiver operating characteristic (ROC) curves, the areas under these curves and confidence intervals for these areas were computed using the Statistical Package for Social Sciences version 10.0 (Statistical Analysis Systems, Chicago, Illinois).

Ten one-hidden-layer feed forward neural networks were designed based on morphological variables and a further two networks were designed using tumour marker variables. A limited number of parameters and hidden nodes were included in the model to avoid overtraining and consequent bad generalisation on new examples. The learning rule was Multilayer Normal Feed forward and the error calculations were based on mean square error. As this is a scale variant technique, the inputs were scaled prior to input. Her-2/neu was multiplied by 0.1; age, CA 15-3 and CA 72-4 were multiplied by 0.01 and CA 125 and CA 19-9 were multiplied by 0.001. To bias the network performance towards accurate detection of malignant masses, those inputs were given a double weight. Training consisted of repeating iterations of the network equations until the mean square error between the desired output and actual output of the network was minimised. Overtraining was avoided by stopping training once test set error began to increase after the initial decrease. The network had the function1/(1+exp(-z))

as a saturation characteristic where exp() is the natural exponent and z is the weighted sum of the input variables for the node.

The models were then validated on the test set of cases. A manual selection of the best performing network was based on the area under the ROC curve and specificity and sensitivity for both the training and test sets. A manual selection of the best performing network was based on the area under the ROC curve and specificity and sensitivity for both the training and test sets.

2.12 Ethics Committee approval

The Local Research and Ethics Committee of Kings College Hospital (00-048) granted approval for all ultrasound examinations, venesection and assays performed in 2000. All patients were given a patient information leaflet and informed consent was taken before inclusion in the study. The patient's General Practitioner (GP) was informed of the trial by letter. The patient information leaflet, consent form and GP letter are included in the Appendices II to IV.

Chapter 3

The use of novel tumour markers to predict the presence of malignancy

3.1 Background

CA 125 has long been established as a discriminatory variable in the pre-operative diagnosis of ovarian cancer (Bast 1983). More recently, other tumour markers have been identified in the serum of women with ovarian cancer. This study investigates the use of four tumour markers: CA 19-9, CA 72-4, CA 15-3 and Her-2/neu.

CA 19-9 is a mucin-type glycoprotein with a molecular weight of approximately ten kiloDaltons. It is expressed in biliary and pancreatic tissues and is thought to aid metastasis through its action as an anti-adhesion molecule. Serum CA 19-9 is raised in 29 to 57% of primary ovarian tumours (Kudoh 1999, Gadducci 1992, Rosen 2005). It is commonly expressed by mucinous tumours and is elevated in 76.9% of mucinous ovarian cystadenocarcinomas (Kudoh 1999).

CA 15-3 is expressed by the MUC1 gene in patients with breast, ovarian and lung carcinoma and shed into the sera. It is a heterogeneous high molecular weight mucin and produced by secretory epithelium. MUC1 has an inhibitory role in cell to stroma interaction, hypothesised to be a key factor in the detachment of cells from stroma allowing for the dissection of the connective tissue and enabling the spread of cells (Nassar 2004). Immunostaining studies have shown CA 15-3 to be present in papillary serous and endometrioid cancers but absent in ovarian epithelium and inclusion cysts (Drapkin, 2004). Work by Feng found that over-expression of CA 15-3 was associated with a higher histological grade and stage than other tumours (Feng 2002). Serum CA 15-3 is elevated in 56% of women with

ovarian cancer (Devine, 1994) CA 15-3 diagnoses ovarian malignancy with a sensitivity of 57.1% and a specificity of 93.9% (Gadducci, 1992).

Serum CA 72-4 is a high molecular weight glycoprotein that is raised in the presence of gastrointestinal and ovarian cancer. It is elevated in 50 to 66% of ovarian cancer patients (Gero 1989, Scambia 1990, Udagawa 1998). CA 72-4 is expressed in few normal tissues such as endometrium and transitional colonic mucosa. Serum levels in healthy volunteers are low, giving CA 72-4 a high specificity in the diagnosis of malignancy (Woolas 1995, Hasholmer 1986, Gero 1989).

Her-2/neu is over expressed in 30 to 32% of ovarian cancers on tissue staining (Slamon 1989, Berchuk 1990). It encodes a transmembrane growth factor receptor with a molecular weight of 185 000 Daltons. The extracellular domain is shed into serum and is elevated in 15% of women with ovarian cancer (Mackenzie, 1993). It confers a poor prognosis for grade and stage (Slamon 1989, Berchuck 1990, Meden 1997). Over-expression of HER-2/*neu* induces endothelial cell retraction, enhancing the metastatic potential of the tumour cells (Carter, 2001).

Several groups have combined tumour markers to improve pre-operative diagnosis of ovarian cancer. Woolas measured eight different markers in the serum of women with a pelvic mass of varying histology: 45% of the masses were malignant, including ovarian cancer, endometrial cancer, sarcoma, colorectal cancer and squamous cell cancer. He used a complex decision tree with twenty terminal nodes and inputs of CA 125, OVX1, lipid-associated sialic acid (LASA), CA 15-3 and CA 72-4. This gave a sensitivity of 82.3 with a specificity of 93.2% (Woolas 1995). In a further study on the same data set, Zhang developed a neural network using CA 125, CA 15-3, CA 72-4 and

LASA to diagnose malignancy. They achieved a sensitivity of 79% in the diagnosis of malignancy with a specificity of 87.5%. They compared the results to CA 125 alone, which achieved a higher sensitivity (82.4% with a lower specificity (68.4%).

In an attempt to improve ovarian cancer screening Skates measured CA 125 II, CA 15-3, CA 72-4 and macrophage-colony stimulating factor (M-CSF) in a population of healthy controls and women with ovarian cancer. They found for a set specificity of 98%, a combination of CA 72-4, M-CSF and CA 125 gave a sensitivity of 85% with the use of classification trees, logistic regression analysis and mixture discriminant analysis (Skates, 2004).

Few groups have examined tumour markers in combination with other variables for the preoperative diagnosis of ovarian cancer. Schutter combined an ultrasound scoring system, physical examination, and CA 125 and CA 72-4 values to diagnose malignancy. CA 72-4 alone at a cut-off of 3U/ml gave a sensitivity of 61% with a specificity of 93%. The addition of CA 72-4 to the logistic model improved the diagnostic accuracy from 81 to 87% (Schutter 1997).

The aim of this study was to evaluate the diagnostic potential of tumour markers in women with adnexal masses and to develop new models for the preoperative diagnosis of ovarian cancer.

3.2 Methods

Demographic and ultrasound variables were measured and recorded as described in Sections 2.3 and 2.4.

3.2.1 Venous blood samples

Blood samples were taken with informed consent from a peripheral vein. The blood was centrifuged within 2 hours for 15 minutes at 300 rpm and aliquots of the supernatant stored at –20^{0}C until analysis. Each aliquot was thawed only once with any remaining serum discarded after thawing.

3.2.2 Measurement of tumour markers

All assays were performed in duplicate and any samples where either of the duplicate values did not fall into within 10% of the other were reassayed. When a quality control value differed by greater than 10% from previous assay means, the assay was repeated. Assays were performed as described in Section 2.6.

3.2.3 Artificial neural networks

The data set was randomly split into two sets for the development of the neural networks; a training set consisting of approximately two third of the cases and a test set containing the remainder of the cases. The variables used

for classification were age, CA 125, CA 15-3, CA 72-4, CA 19-9 and Her-2/neu. Forward stepwise logistical regression was used to identify the most significant variables in the training set. This process of randomly splitting the data set and the application of logistic regression was repeated a total of seven times. The desired network output was coded as '0' for benign and '1' for malignant masses. Seven different neural networks were developed.

Receiver operator characteristic curves were constructed for each of the models and the diagnostic performance of the models was compared using the areas under the curve.

3.3 Results

One hundred and thirty-five women were included in the study. Their mean age was 45 with an interquartile range of 33 to 54 years. All women had blood assayed for CA15.3, CA19.9, CA72.4, CA 125 and Her-2/neu.

3.3.1 Histology

Twenty of the masses were invasive; twelve were borderline with one hundred and two benign masses. The histological classification is shown in Table 3.1.

Table 3.1: Histology of masses

Histological Type	Nature	Number of masses
Endometrioma	Benign	23
Dermoid cyst	Benign	26
Serous cystadenoma	Benign	13
Mucinous cystadenoma	Benign	14
Fibroma	Benign	2
Simple ovarian cyst	Benign	2
Peritoneal pseudocyst	Benign	1

Histological Type	Nature	Number of masses
Tubal	Benign	6
Cystadenofibroma	Benign	12
Miscellaneous	Benign	4*
Serous cystadenoma	Borderline	6
Mucinous cystadenoma	Borderline	5
Cystadenofibroma	Borderline	1
Serous cystadenocarcinoma	Invasive	5
Mucinous cystadenocarcinoma	Invasive	3
Endometrioid adenocarcinoma	Invasive	5
Granulosa cell Tumour	Invasive	1
Clear cell carcinoma	Invasive	2
Undifferentiated carcinoma	Invasive	4

* Includes 2 tubo-ovarian abscesses, 1 fibroid and 1 thecoma

3.3.2 Tumour markers

CA 125 was higher in the borderline and invasive groups of masses than in the benign group ($p<0.001$, Kruskal Wallis test) (Table 3.2). It was also significantly higher in the malignant group than the benign group ($p<0.001$,

Mann-Whitney U test).

CA 72-4 was also significantly higher in the invasive than benign and borderline groups (p=0.005, Kruskal Wallis test) and in the malignant as compared to the benign group (p=0.002, Mann-Whitney U test).

CA 15-3 was higher in the invasive than the borderline and benign groups (p=0.010, Kruskal Wallis test). It was also higher in the malignant group than the benign group (p=0.003, Mann-Whitney U test).

There was no significant difference between the groups with CA 19-9 (p=0.26 Mann-Whitney U test) or Her-2/neu. The distribution of tumour markers in benign, borderline and invasive masses are shown in Figures 3.1-5.

Table 3.2: Values of CA 125, CA 72-4, CA 15-3, CA 19-9 and Her-2/neu with different histology. Values are mean with standard deviations in brackets.

	Benign	Borderline	Invasive	All Malignant
CA 125	37.8 (58.0)	144.8 (173.2)	1420.9 (2641.7)	982.3 (2211.6)
CA 72-4	1.08 (1.81)	5.31 (12.1)	61.5 (118.3)	42.2 (99.1)
CA 15-3	15.4 (6.50)	37.7 (57.6)	66.5 (89.2)	56.6 (80.0)
CA 19-9	17.9 (37.3)	2486.7 (6903.5)	1543.1 (4369.9)	1867.5 (5282.1)
Her-2/neu	8.9 (1.9)	9.0 (2.3)	9.8 (2.5)	9.5 (2.4)

Figure 3.1: Distribution of CA 125 with benign, borderline and invasive masses

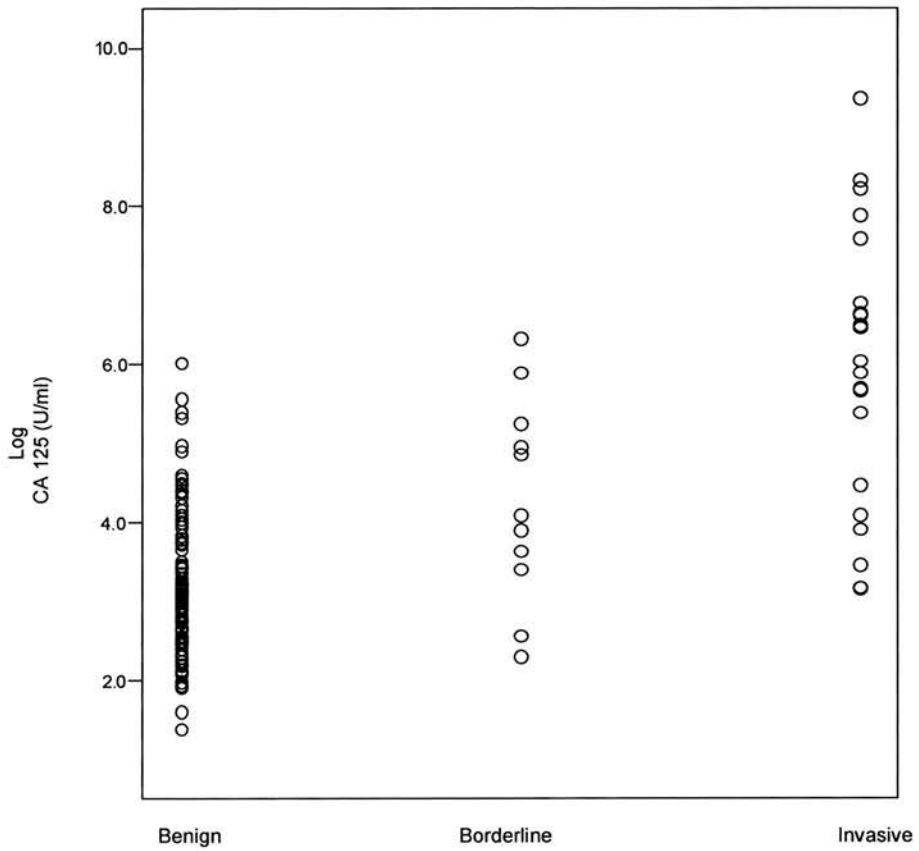


Figure 3.2: Distribution of CA 72-4 with benign, borderline and invasive masses

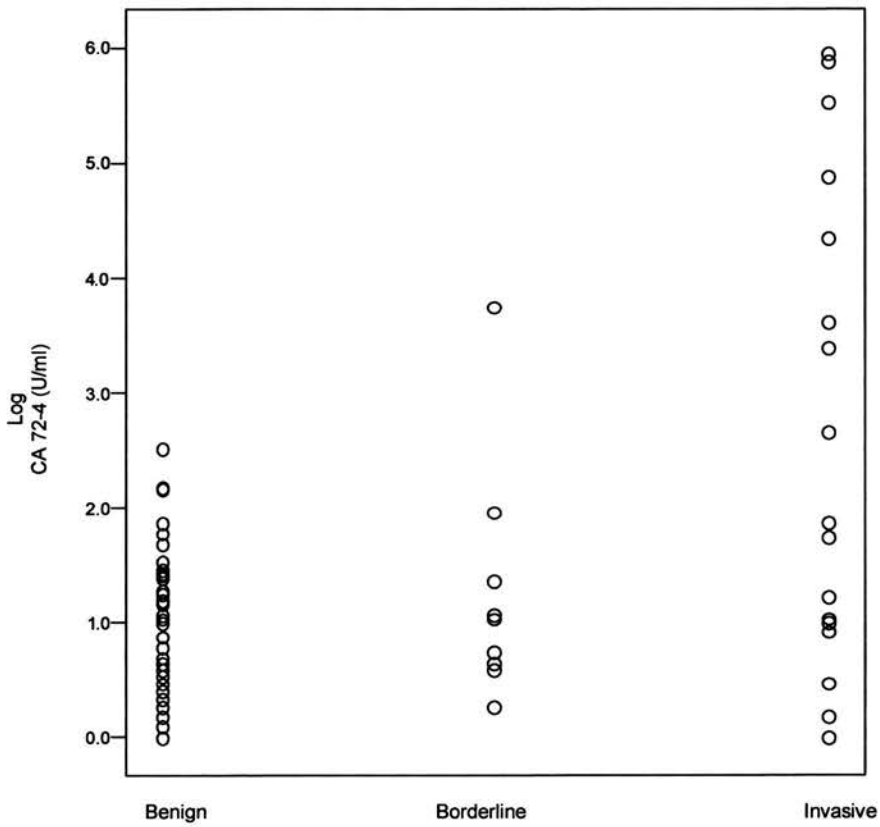


Figure 3.3: Distribution of CA 15-3 with benign, borderline and invasive masses

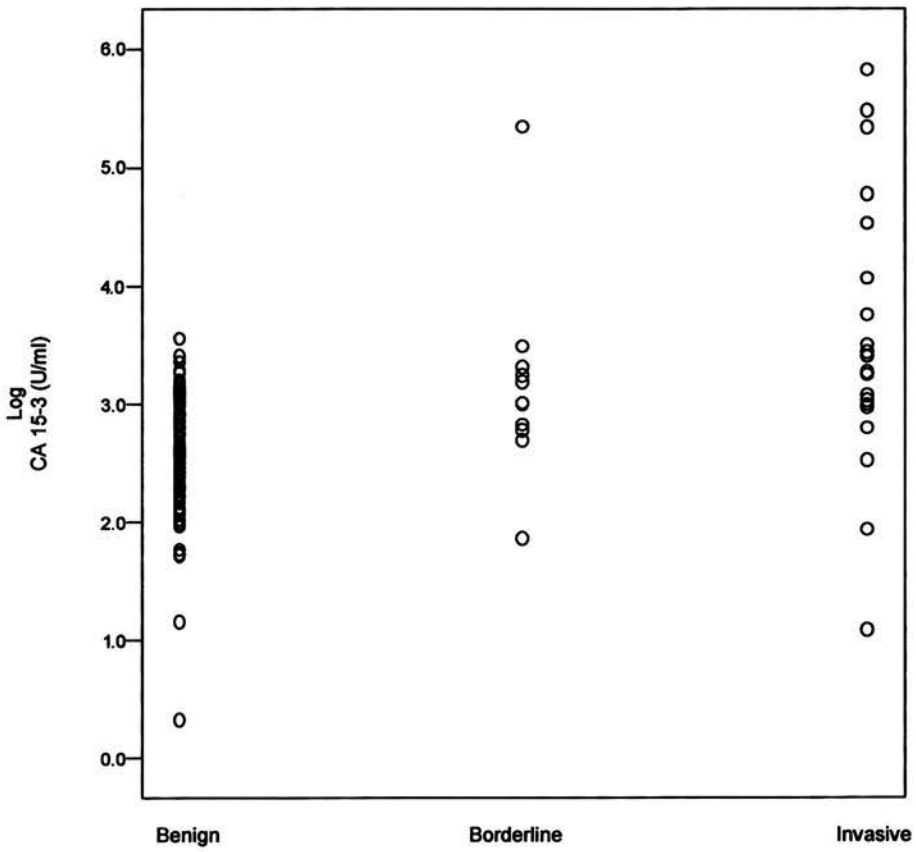


Figure 3.4: Distribution of CA 19-9 with benign, borderline and invasive masses

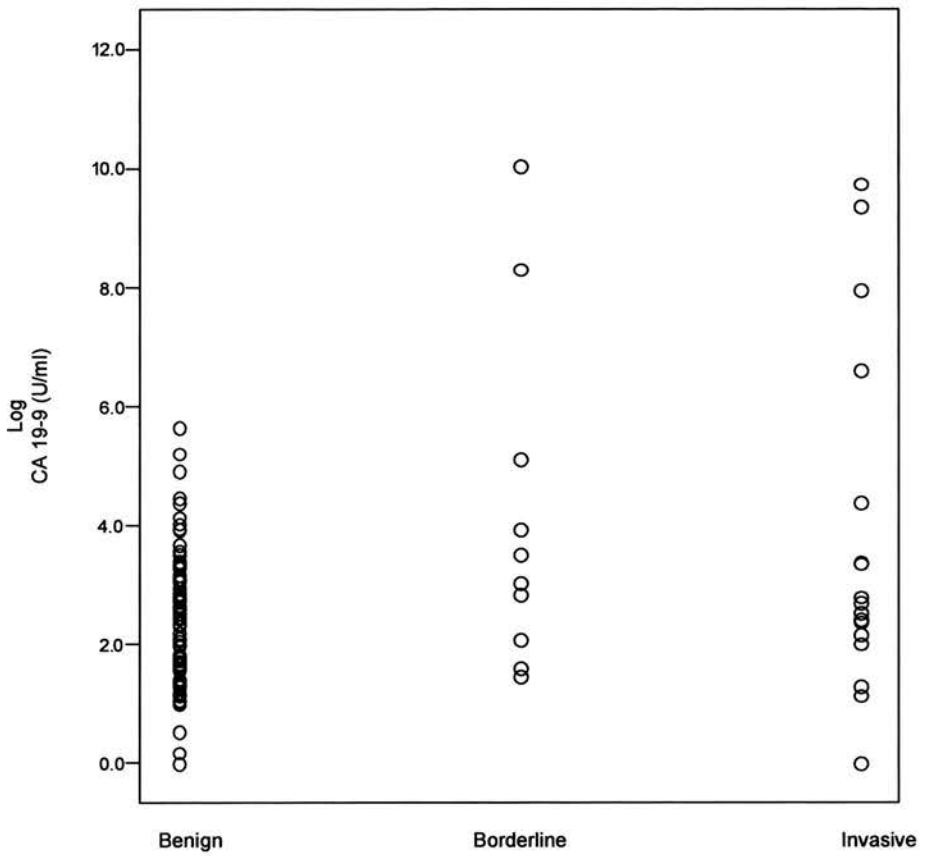
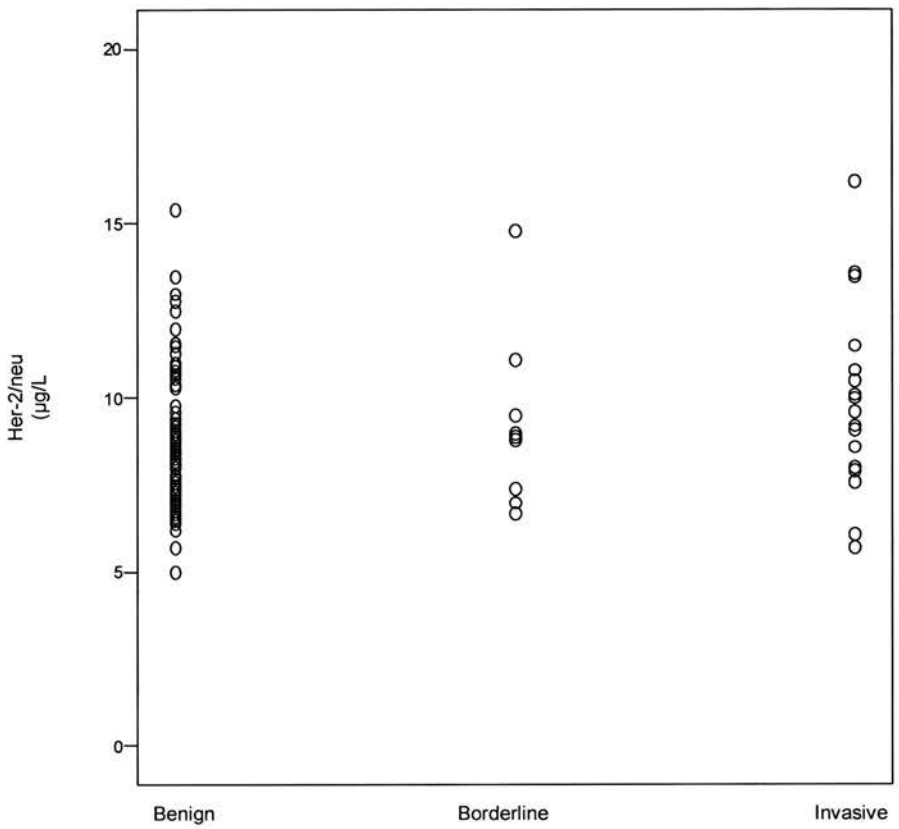


Figure 3.5: Distribution of Her-2/neu with benign, borderline and invasive masses



Comparison of the ROC curves (Figure 3.6 and Table 3.3) showed CA 125 to be the best discriminator between benign and malignant masses. Age and CA 15-3 had a similar ability to differentiate between the groups. CA 125 was a significantly better discriminator than CA 19-9, CA 72-4 and Her-2/neu (using the methods of Hanley and McNeill to compare the AUC).

Figure 3.6: Receiver Operator Characteristic curve for tumour markers and age against benign and malignant histology

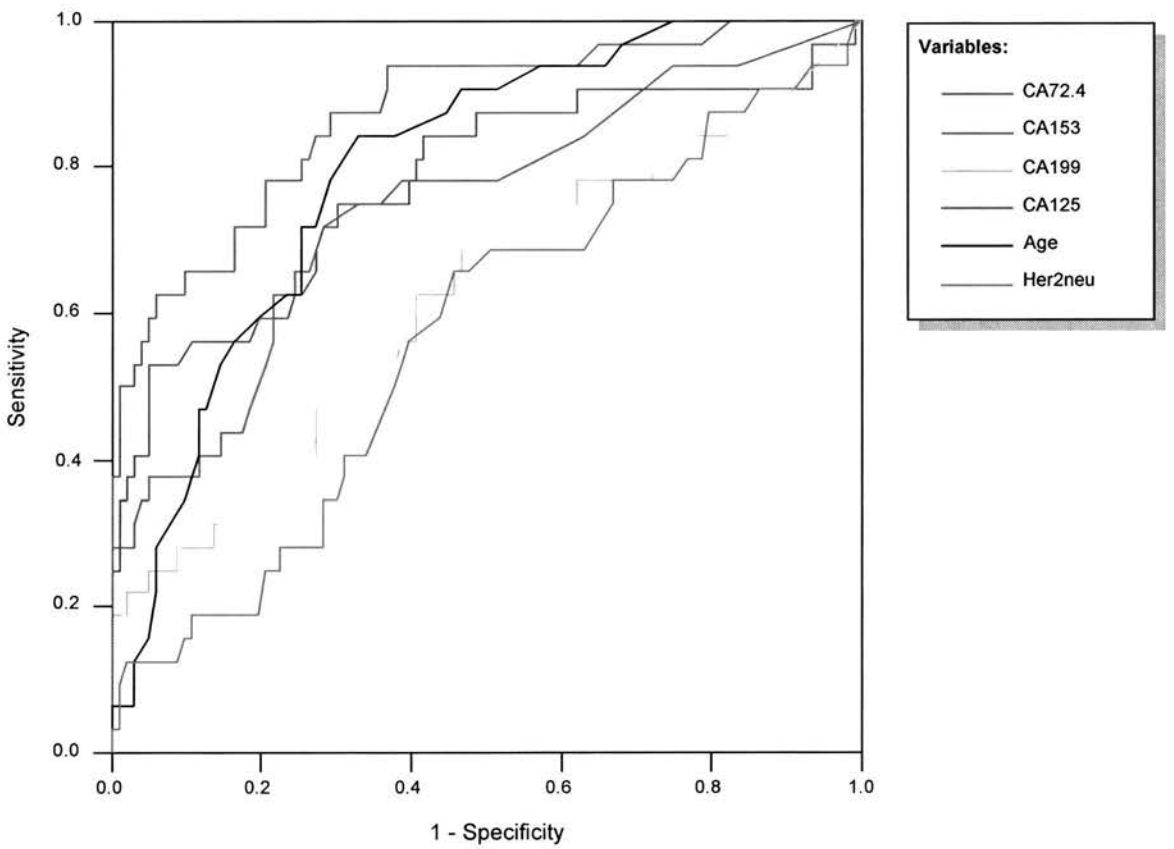


Table 3.3: Area under the ROC curve for tumour markers and age.

Variable	Area under curve	95% Confidence interval
CA 125	0.874	0.801-0.947
CA 72-4	0.747	0.643-0.851
CA 15-3	0.778	0.673-0.884
CA 19-9	0.628	0.507-0.748
Her-2/neu	0.571	0.455-0.686
Age	0.797	0.717-0.878

3.3.2.1 Tumour markers in benign disease

CA 125 differed significantly ($p<0.001$, Kruskall Wallis test) between the four groups of benign tumours: dermoid, endometrioma, serous cystadenoma and cystadenofibroma, and mucinous cystadenoma (Table 3.4). CA 72-4 was also significantly different between the groups ($p=0.001$, Kruskall Wallis test). Pairwise comparisons between the groups were carried out for CA 125 and CA 72-4. A Mann Whitney U test was used to compare the groups with a Bonferroni adjustment for significance. Both tumour markers were higher in endometriomas than dermoids ($p<0.01$). They were also both higher in endometriomas than in the serous cystadenoma group ($p<0.01$). CA 125 was higher in endometriomas than in mucinous cystadenomas ($p=0.048$). CA 125 was higher in mucinous cystadenomas than in the serous cystadenoma group although not significantly so ($p=0.09$). There was no difference between either dermoids and serous cystadenomas or dermoids and mucinous cystadenomas for CA 72-4 or for CA 125.

Her-2/neu, CA 15-3 and CA 19-9 values did not differ between the four benign groups.

Table 3.4: Values of CA 125, CA 72-4, CA 15-3, CA 19-9 and Her-2/neu with different benign histologies. Values are mean with standard deviations in brackets.

	Dermoid	Endometrioma	Serous cystadenoma and cystadenofibroma	Mucinous cystadenoma
n	26	23	19	14
CA 125	22.5 (19.9)	82.5 (97.3)	12.9 (9.3)	26.1 (23.5)
CA 72-4	1.1 (2.4)	1.5 (1.7)	0.658 (1.26)	1.5 (2.3)
CA 15-3	15.1 (6.3)	18.0 (6.5)	13.7 (5.7)	14.3 (6.3)
CA 19-9	21.9 (31.7)	24.3 (39.3)	6.7 (10.3)	28.0 (74.0)
Her-2/neu	8.9 (1.9)	8.4 (1.4)	8.6 (1.2)	8.5 (1.8)

3.3.2.2 Tumour markers in invasive disease

CA 125 levels were higher in endometrioid, clear cell, serous and undifferentiated cancers than in mucinous cystadenocarcinomas (Table 3.5). CA 72-4 and CA 19-9 were higher in mucinous and endometrioid cancers. CA 15-3 was highest in endometrioid and serous cancers. None of these trends were statistically significant due to the small numbers in each group.

Table 3.5: Values of CA 125, CA 72-4, CA 15-3, CA 19-9 and Her-2/neu with different invasive histologies. Values are mean with standard deviations in brackets.

	Serous	Mucinous	Clear cell	Endometrioid	Undifferentiated
n	5	3	2	4	4
CA 125	1086.6 (1170.7)	334.6 (293.5)	2186.4 (2781.2)	3196.6 (5653.8)	1173.5 (1764.9)
CA 72-4	47.1 55.8)	119.3 (205.0)	2.5 (3.2)	72.2 (119.6)	100.5 (187.0)
CA 15-3	74.5 (87.4)	26.5 (6.8)	56.6 (51.6)	106.6 (157.1)	27.1 (13.0)
CA 19-9	5.3 (6.1)	6609.9 (9083.8)	385.1 (504.7)	2916.4 (5809.1)	6.3 (7.4)
Her-2/neu	8.2 (2.3)	9.2 (1.3)	10.4 (1.6)	12.5 (3.1)	10.6 (2.9)

3.3.3 Decision tree analysis

Two types of decision trees were designed, the first to discriminate between benign and malignant adnexal masses, the second to discriminate between benign, borderline and invasive masses. Data variables entered into the models were age, CA 125, CA 15-3, CA 19-9, CA 72-4 and Her-2/neu.

The first model retained two variables, CA 125 and CA 15-3, after pruning.

The decision rules were:

CA 125 > 278 denotes a malignant mass

CA 15-3 > 61.69 denotes a malignant mass

The model comprised five nodes with two levels as shown in Figure 7. The model had three terminal nodes: node 3 contained the cases classified as benign and nodes 2 and 4 contained the cases classified as malignant. This model gave a sensitivity of 56.3% with a specificity of 98.1% for the diagnosis of malignant masses.

The second model again retained the same variables CA 125 and CA 15-3. The structure of the model was similar with five nodes and two levels. The three terminal nodes classified the cases into benign (node 3), borderline (node 4) and invasive (node 2). This model did not perform well in the classification of borderline tumours with a sensitivity of 9.1% and a specificity of 98.4%. However, the detection of invasive malignancy had a sensitivity of 67% with a specificity of 97.3%. This gave a weighted kappa value of 0.654 for the agreement between the decision tree and histological outcome. A weighted kappa between 0.61 and 0.80 suggests good agreement (Altman).

Figure 3.7: Decision tree 1 for the discrimination between benign and malignant masses. The percentages of benign and malignant are shown with the absolute numbers in each node.

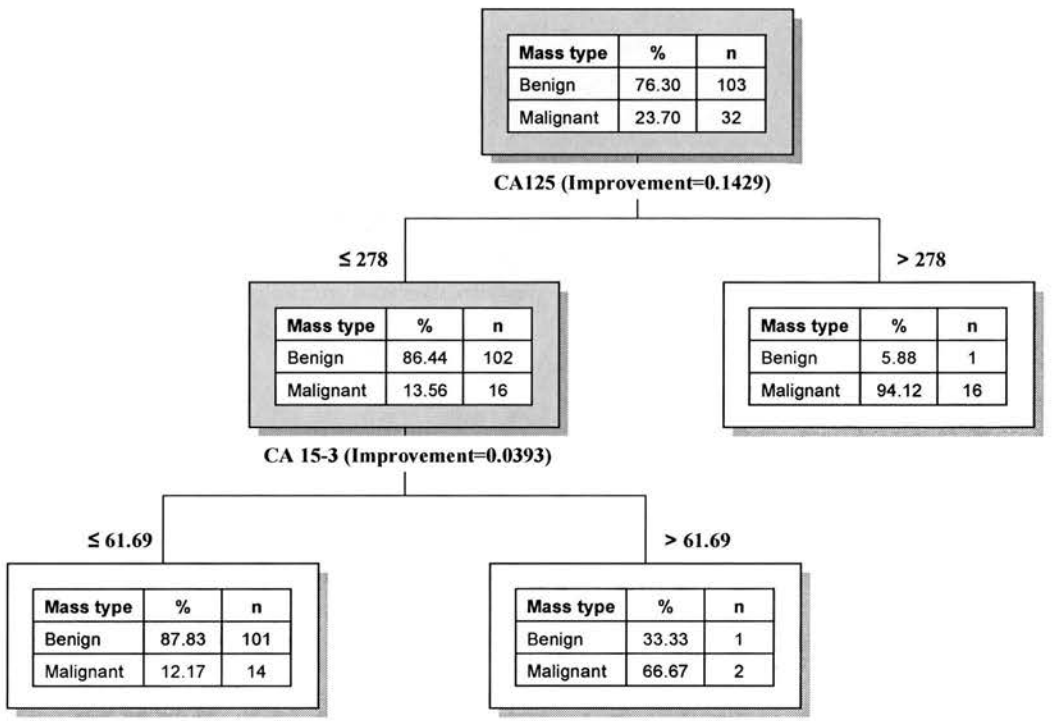
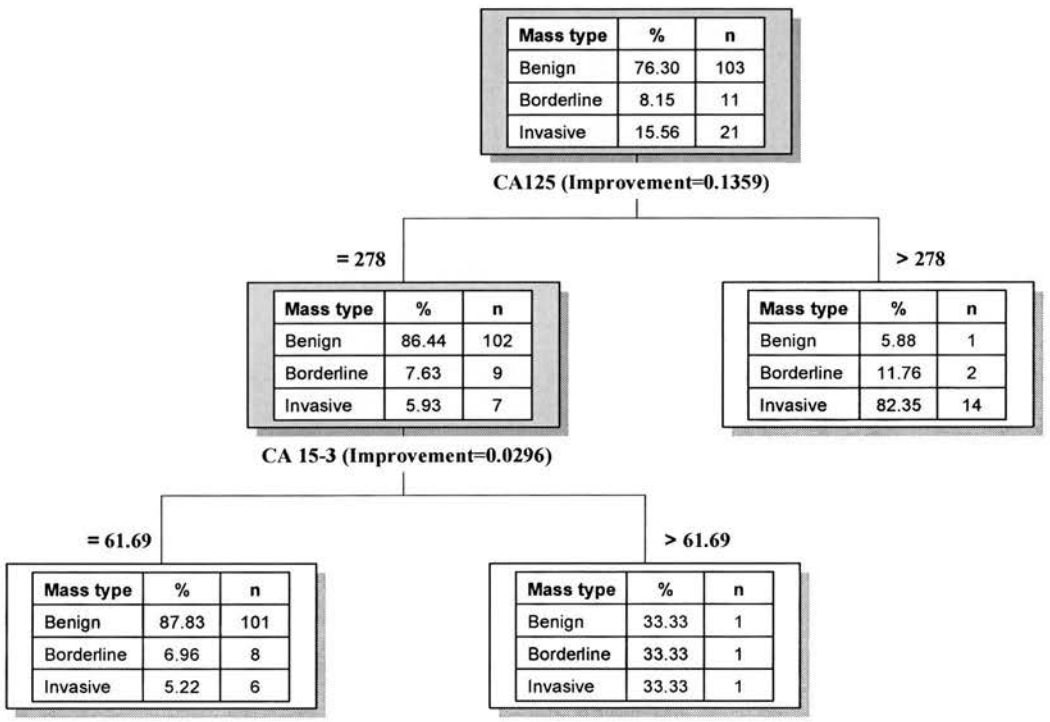


Figure 3.8: Decision tree 2 for the discrimination between benign, borderline and invasive masses. The percentages of benign and malignant are shown with the absolute numbers in each node.



3.3.4 Artificial neural networks

The inputs for all the networks were selected using logistic regression from age, CA 125, CA 19-9, CA 72-4, CA 15-3 and Her-2/neu. The inputs were pre-processed to standardise the size of the inputs to the network: Her-2/neu was divided by 10; age, CA 15-3 and CA 72-4 were divided by 100; CA 125 and CA 19-9 were divided by 1000.

The data set was split into randomly selected training and test sets. The most significant variables were then selected using forward stepwise logistic regression. The data was then split six further times. All the data splits

except one yielded the same input variables. Six splits gave age, CA 15-3 and CA 125. The other split gave CA 72-4 in addition. Two types of neural networks were developed: networks with three input variables (age, CA 125 and CA 15-3) and networks with four inputs (age, CA 125, CA 15-3 and CA 72-4).

The networks were developed by changing the network architecture based on trial and error. Twelve models were trained and cross-validated using the test set of patient records. The best network was manually selected using the sensitivity and specificity values, together with the area under the ROC curve, using both the training and test set.

The first network (TNN1) contained a single hidden layer with three input nodes, two hidden nodes and one output node. Training was stopped after 654 iterations when the training error was 0.283 and the test error 0.3701.

The mathematical equation of TNN1 is:

$$\mathbf{Y = 1/(1+\exp(z_y))\ \ with\ z_y = -2.6988h_1 - 2.8536h_2 + 1.7429}$$

Where:

$$\mathbf{h_1 = 1/(1+(\exp z_1))\ with\ z_1 = -2.8428(age) - 1.2728(CA\ 15\text{-}3) - 4.5840(CA\ 125) + 2.2115}$$

$$\mathbf{h_2 = 1/(1+(\exp z_2))\ with\ z_2 = -1.6406(age) - 1.3794(CA\ 15\text{-}3) - 4.8990(CA\ 125) + 1.6554}$$

With the probability of malignancy set at 15%, TNN1 gave a sensitivity of 84.2% and a specificity of 79.16% on the training set. When cross validated on the test set, this decision level gave a sensitivity and specificity of 100%

and 90.9% respectively. An ROC curve constructed on the training, test and complete data set gave an area under the curve of 0.911; 0.959 and 0.932 respectively (Figure 10).

The second network (TNN2) had four selected inputs: age, CA 125, CA 72-4 and CA 15-3. The inputs were pre-processed to standardise the size of the inputs to the network: age, CA 15-3 and CA 72-4 were divided by 100, and CA 125 was divided by 1000. The network consisted of four input nodes, two hidden nodes and one output node. Training was stopped at 540 iterations when the test error was 0.3662 with a training error of 0.2794.

The mathematical formula of the network is:

$$\mathbf{Y = 1/(1+\exp(z_y))\ \ with\ z_y = -1.9169h_1 - 3.3743h_2 + 1.7135}$$

Where:

$$\mathbf{h_1 = 1/(1+(\exp z_1))\ with\ z_1 = -2.2524(age) - 1.3998(CA\ 72\text{-}4) - 2.7744(CA\ 15\text{-}3) - 4.8949(CA\ 125) + 3.7748}$$

$$\mathbf{h_2 = 1/(1+(\exp z_2))\ with\ z_2 = -2.0157(age) - 0.9519(CA\ 72\text{-}4) - 1.5292(CA\ 15\text{-}3) - 3.3407(CA\ 125) + 2.4497}$$

At a probability of malignancy set at 20%, the second neural network gave a sensitivity of 88.9% and a specificity of 81.8% on the training set. Cross validation on the test set gave a sensitivity of 60% with a specificity of 90.9%. An ROC curve constructed on the training, test and complete data set gave an area under the curve of 93.2%, 85% and 89.9%. The ROC curve for the whole data set is shown in Figure 3.9 below.

Figure 3.9: Receiver Operator Characteristic curve for the two neural networks

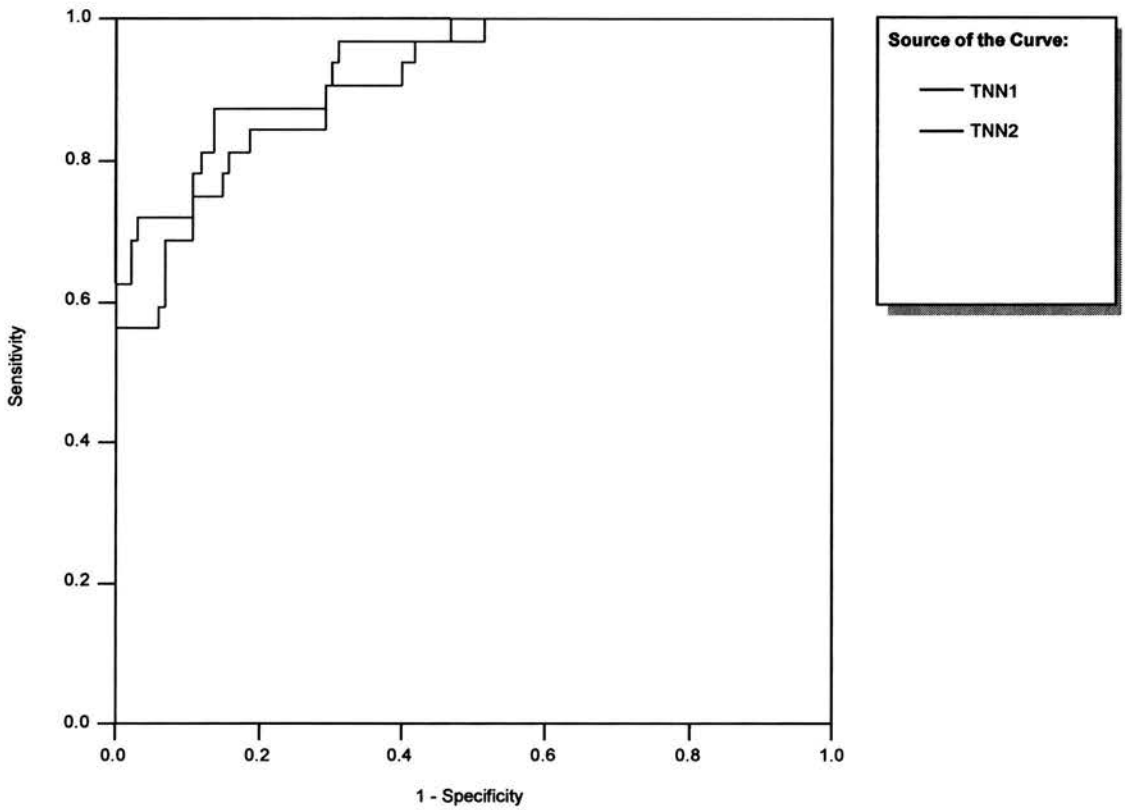


Table 3.6: Comparison of the performance of TNN1 and TNN2

There was no significant difference between the AUC for the complete sets of the two networks using the method described by Hanley and McNeil (1983).

Network	Data Set	n	Probability of malignancy (%)	Sensitivity (%)	Specificity (%)	Area under ROC curve (%)	95% CI
TNN1	Training	91	15	84.2	79.2	0.911	0.813, 1.0
	Test	42	15	100	90.9	0.959	0.893, 1.00
	Complete	135	15	-	-	0.932	0.874, 0.99
TNN2	Training	91	20	88.9	81.8	0.932	0.874, 0.99
	Test	42	20	60	90.9	0.85	0.71, 0.99
	Complete	135	20	-	-	0.899	0.835, 0.963

3.4 Discussion

The tumour markers CA 125, CA 15-3 and CA 72-4 were all significantly related to the histological outcome. All three tumour markers were higher in borderline than benign masses and in invasive than borderline masses. Her-2/neu was not significantly different between benign, borderline and invasive masses. CA 19-9 was higher in the borderline group than the benign and invasive groups, although this failed to reach significance. This is consistent with the findings of other workers who have found a proportionally higher CA 19-9 than CA 125 in borderline tumours (Engelen et al, 2000).

Of all the markers, CA 125 was the most significantly different between the benign and malignant groups and the benign, borderline and invasive groups. The mean level of CA 125 in borderline tumours was 144 kU/l. This was higher than the level of 44kU/l found in a previous study (Aslam, 2000). Both CA 15-3 and CA 72-4 were more significantly different between benign and malignant than between benign, borderline and invasive masses.

The best univariate discrimination between benign and malignant masses was seen with CA 125, age, CA 15-3 and CA 72-4 on comparison of the ROC curves. The AUC for CA 125 was 0.874 and 0.747 for CA 72-4. This is an improved performance on previous work, which found the AUC to be 0.69 for CA 125 and 0.691 for CA 72-4 (Udagawa, 1998).

CA 125 was significantly elevated in endometriomas as compared to dermoid cysts and serous cystadenomas. Previous workers have found a raised CA 125 level in endometriosis and a decrease in serum CA 125 with

gonadotrophin-releasing hormone agonist therapy (Ozaksit, 1995). CA 72-4 was also higher in endometriomas than in dermoid cysts and serous cystadenomas. CA 19-9 was not higher in dermoid cysts in this study, in contrast to previous findings (Dede, 2006).

CA 125 was also higher in serous, clear cell and endometrioid cancers and lower in mucinous carcinomas. CA 19-9 and CA 72-4 were raised in mucinous tumours as found in previous studies (Kudoh 1999, Gadducci 1992, Rosen 2005) although this did not reach significance. Her-2/neu was not raised significantly in any of the groups of ovarian cancer. CA 15-3 was raised in serous and endometrioid cancers as found in previous work (Drapkin, 2004).

The decision tree models identified CA 125 and CA 15-3 as the strongest predictors of malignancy amongst the tumour markers. This is consistent with the univariate marker analysis, which showed CA 125 to be the most significant predictor of malignancy (AUC 0.874) with CA 15-3 as the second strongest predictor (AUC 0.778). The benign vs malignant decision tree gave a good specificity of 98% but a less impressive sensitivity of 56.3%. The second decision tree gave an improved performance on the detection of malignancy with a sensitivity of 67% whilst maintaining a good specificity of 97.3%. It was less good at detecting borderline tumours with a low sensitivity of 9.1%. Previous studies have also found borderline tumours difficult to classify (Aslam, 2000, Valentin, 2001).

Overall, the second model gave good agreement between prediction and outcome with a kappa value of 0.654. The most predictive cut off levels in this study are similar to the initial cut offs levels used in Woolas' decision tree. That group used a CA 125 of 101.55 with a CA 15-3 of 64.9 kU/l. With

the use of five tumour markers and a complex model with 41 nodes, his group found a sensitivity of 82.3% with a specificity of 93.2% (Woolas 1995). The second decision tree in this study gave an improved specificity but the sensitivity was lower at 67%. This was a much simpler model with only four nodes and, due to its simplicity, is likely to generalise better to new populations.

The neural networks performed significantly better than the decision trees. Their non-linear decision boundaries allowed a better fit to the complex data set. The optimal design of a network including the number of input and hidden nodes is a difficult task which is only answered by the trial and training of a large number of networks to identify the best among them. A large network with a high ratio of nodes to case examples will fit the data well but is likely to generalise poorly when applied to new populations. However, an overly simple network will not be able to retain sufficient information in the weights between the nodes for effective discrimination.

These neural networks performed better than the previously published network (Zhang 1999). This may be due to a number of factors. The network of Zhang was complex with 30 nodes with 144 internodal connections. The models in this study were simpler: TNN1 has six nodes and eight internodal connections; TNN2 has seven nodes with 10 internodal connections. As described above, this avoids overfitting of the data and thus generalises better when confronted with new cases. Zhang et al tested their network on a heterogeneous collection of pelvic masses including 44 non-ovarian malignancies (e.g. lymphoma, colon carcinoma and lymphoma). As the network was only trained on examples of ovarian cancer, poor generalisation in the diagnosis of non-ovarian cancer may be expected.

The results of TNN1 are hopeful but need to be assessed by prospective validation in a new population. Previous publications have highlighted the problem of local minima (Timmerman, 1999b). This is the difficulty of assessing the network performance using mean square error during training. Training is stopped when the mean square error is minimised but due to the non-linear nature of the neurones, further training may lead to a decrease beyond that in the mean square error.

A combination of tumour markers has been shown to provide an improved preoperative diagnosis of ovarian cancer. The use of tumour markers is at least as good as that of morphological models. The application of neural networks to the tumour marker data has provided encouraging results. These results need to be validated in a well-designed prospective trial before they can be used in clinical practice.

Chapter 4

A prospective evaluation of models for the diagnosis of ovarian cancer

4.1 Background

The diagnostic performance of published models in the diagnosis of ovarian cancer varies significantly with reported sensitivities ranging from 0.85 to 1.00 and specificities from 0.77 to 0.97 (Jacobs 1990, Lerner 1994, Tailor 1997, Timmerman 1999, Timmerman 2006). This variation is likely to represent differences in the models, study population and ultrasound technique between centres. All the models were designed based on retrospective data analysis and therefore the results cannot be reliably applied to new populations. Four models have been tested prospectively in different centres on new datasets to validate their performance: Jacobs, Tingulstad, Tailor and Timmerman's logistic regression model.

The aim of this study is to validate the performance of models that have not been tested prospectively and to compare their performance to previously validated models. This is the first clinical validation of the IOTA logistic regression model. Timmerman's neural network has been tested prospectively (Mol, 2001) but significant queries have been raised regarding the study methodology.

4.2 Population

One hundred and seventy six consecutive women referred to King's College Early Pregnancy and Gynaecology Assessment Unit with a clinical suspicion of a pelvic mass or found on scanning to have a mass were recruited into the study. All women underwent a transvaginal ultrasound scan with colour Doppler imaging and CA 125 estimation as part of the clinical investigation of the pelvic mass.

4.2.1 Inclusion Criteria

The inclusion criteria were the presence of a pelvic mass that appeared to originate from the adnexa on ultrasonography and patient consent to inclusion in the study.

4.2.2 Exclusion Criteria

The exclusion criteria were: pregnancy; refusal of transvaginal ultrasonography; lack of fitness for, or refusal of, surgery; previous bilateral oophorectomy or the finding of a benign mass in one ovary and a borderline or invasive mass in the other. In the case of bilateral tumours, the mass with the more suspicious morphological features, as judged by the ultrasonographer, was included. In the case of a similar appearance of both masses, the larger was documented. Informed consent was obtained prior to ultrasound examination and venepuncture.

4.3 Statistical analysis

A database file was set up using Microsoft Excel for Windows (Redmond, WA, USA) to facilitate data entry and retrieval. Statistical analysis was performed using SPSS for Windows (Version 10.0; SPSS Inc, Chicago, IL, USA). Variables were assessed for normality using Q-Q plots. The normally distributed continuous variables were compared to the presence of malignancy using the independent paired samples t test. The nonparametric variables were compared to the histology using the Mann-Whitney test. Nominal variables were compared to the histology using Fisher's exact test.

The data set was randomly split into two sets for the development of the neural networks; a training set consisting of approximately two thirds of the cases and a test set containing the remaining one third. Forward stepwise logistical regression was used to identify the most significant variables in the training set. As there was a large amount of variability in the histological types of masses included in the data set, the random splitting of the data set and logistical regression were repeated ten times. The desired network output was coded as '0' for benign and '1' for malignant masses. Ten different neural networks were developed.

Receiver operator characteristic curves were constructed for each of the models and the diagnostic performance of the models was compared using the areas under the curve.

4.4 Results

One hundred and seventy six women were included in the study. Six women were excluded from data analysis (3.4%): five women declined surgical excision of the mass and one woman had a borderline tumour of the left ovary and a benign cyst of the right. Of the 170 women, 137 had benign masses, 12 were borderline and 21 were invasive malignant tumours (Table 4.1). The majority of the benign masses comprised endometriomas, dermoid cysts and cystadenomas. Nine rare tumours were included in a miscellaneous group of benign tumours. These were one Brenner tumour, one paramesonephric cyst, four tubo-ovarian abscesses, one benign steroid cell tumour, 1 hydatid of morgagni and 1 benign leydig cell tumour. The borderline tumours were all epithelial in origin. They were staged according to the FIGO guidelines (Appendix I) and all found to be Stage Ia tumours. The invasive tumours were predominantly found to be Stage I or Stage III. Twenty of the twenty-one invasive tumours were epithelial in origin, with one leiomyosarcoma of the ovary.

Table 4.1: Histological classification of the adnexal masses

Histological Type	Nature	Stage	Number in Stage	Total number
Endometrioma	Benign	-	-	21
Dermoid cyst	Benign	-	-	19
Serous cystadenoma	Benign	-	-	15
Mucinous cystadenoma	Benign	-	-	18
Fibroma	Benign	-	-	12
Simple ovarian cyst	Benign	-	-	17
Peritoneal pseudocyst	Benign	-	-	5
Tubal	Benign	-	-	7
Cystadenofibroma	Benign	-	-	14
Miscellaneous	Benign	-	-	9
Serous cystadenoma	Borderline	I	6	6
Mucinous cystadenoma	Borderline	I	5	5
Cystadenofibroma	Borderline	I	1	1
Serous cystadenocarcinoma	Invasive	I	2	
		II	1	
		III	6	

Histological Type	Nature	Stage	Number in Stage	Total number
		IV	1	10
Mucinous cystadenocarcinoma	Invasive	I	2	3
		II	1	
Endometrioid adenocarcinoma	Invasive	I	1	3
		III	2	
Clear cell carcinoma	Invasive	II	1	1
Undifferentiated carcinoma	Invasive	I	2	3
		II	1	
Leiomyosarcoma	Invasive	I	1	1

4.4.1 Demographic variables

86 of the women were Caucasian (50.6%); 63 were black or black British (37.1%); 14 Asian or British Asian (8.2%); 3 Chinese or other ethnic group (1.8%), and 4 were of mixed race (2.4%).

112 of the women were premenopausal (66%). Of the 58 women who were postmenopausal, the mean age for menopause was 4.3 years previously. 17 of the women had previously undergone a hysterectomy.

46 of the women had never been pregnant (27%) and 88 had one or more children (range 1 to 8). 36 women had had one or more miscarriage (range 1 to 6) and 31 women had had one or more terminations of pregnancy (range 1 to 5).

24 of the postmenopausal women were using hormone replacement therapy (42%). 32 of the premenopausal women were using hormonal contraception, 25 were using the combined oral contraceptive pill, 4 intramuscular progesterone injection, 2 the progesterone only pill and 1 the levonorgestrel intrauterine system. 53 of women (31%) had previously taken the combined pill for an average of 9.6 years (range 1 to 26 years).

Seven of the women had a personal history of breast cancer; none had a personal history of ovarian cancer. Seventeen women had a family history of breast cancer and five women of ovarian cancer.

Table 4.2: Comparison of demographic variables to histology

Variable	Benign	Malignant	Mean Difference	95% CI Difference between means	Statistical test	p
Mean age (SD)	42.4 (14.3)	52.4 (19.1)	-9.883	-15.7 to –4.1	t-test	0.007
Years postmenopausal (SD)	11.0 (8.1)	16.7 (12.4)	-5.711	-11.1 to –0.2	t-test	0.41
Menopausal (%)	27.7	57.6	-	-	Chi-Square	0.001
Hysterectomy (%)	8.0	18.0	-	-	Chi-Square	0.155
Mean number of previous pregnancies (range)	2.3 (0–8)	1.9 (0–5)	-	-	Chi-Square	0.31
Hormonal use (%)	35	27.3	-	-	Chi-Square	0.42
Personal history of breast cancer (%)	3.6	6.1	-	-	Chi-Square	0.38
Personal history of ovarian cancer (%)	0	0	-	-	Chi-Square	1.00
Family history of breast cancer (%)	10.2	9.1	-	-	Chi-Square	0.49
Family history of ovarian cancer (%)	2.2	6.1	-	-	Chi-Square	0.41

The presenting symptom leading to the scan referral was most commonly pain, abnormal bleeding or abdominal distension (Table 4.3). Nineteen adnexal masses were incidental findings in women having scans for non-gynaecological reasons.

Seventy-three women presented with pelvic pain. The severity of the pain varied from mild pain to pain severe enough to limit daily activities (Table 4.4). There was no significant difference between the severity of pain in benign and malignant masses. Urinary symptoms and abdominal distension were more commonly associated with invasive and borderline masses (p=0.017).

Table 4.3: Comparison of presenting symptom to histology

Presenting Symptom	Benign (%)	Malignant (%)	Total
Abnormal bleeding	39 (84.8)	7 (15.2)	46
Abdominal distension	16 (64.0)	9 (36.0)	25
Pain	62 (84.9)	11 (15.1)	73
Urinary symptoms	3 (42.9)	4 (57.1)	7
Incidental	17 (80.6)	2 (19.4)	19

Difference between groups p = 0.017 (Fishers exact test)

Table 4.4: Severity of pain

Pain grade	Severity of pain	Benign	Malignant	Total
0	None	74	22	96
1	Mild	40	5	45
2	Requiring analgaesia	4	4	8
3	Limiting activity	19	2	21

4.4.2 Ultrasound variables

All women were examined with ultrasound. Thirty-five women had bilateral masses; the remaining one hundred and thirty five women had a single mass only. The masses predominantly appeared to arise from the ovaries (90%) with 5% of masses of tubal origin. In 2% the origin was uncertain and the mass arose from a structure other than the tube or ovary in 3% of cases.

The size of the masses ranged from 10 mm to 251 mm in longest dimension with a median of 71mm and an interquartile range of 60.75. Their volume was spread between 10 mm^3 to 3187 mm^3 with a median of 101 mm^3 and an interquartile range of 346.5 (Table 4.5).

Table 4.5: Distribution of tumour size/mm and volume/mm^3 in benign and malignant tumours

Variable	Nature	Mean	Standard deviation	Test of significance	p
Tumour diameter 1	Benign	77.4	44.7	t-test	0.002
	Malignant	110.7	48.6		
Tumour diameter 2	Benign	63.6	34.5	t-test	0.002
	Malignant	89.8	42.1		
Tumour diameter 3	Benign	51.7	28.6	t-test	0.002
	Malignant	73.6	35.3		
Volume	Benign	270.7	497.8	Mann-Whitney	0.001
	Malignant	605.9	628.0		

The majority of the masses were unilocular, multilocular or multilocular solid with only four cases being unclassifiable (Table 4.6). There was a significant difference between the distribution of benign and malignant masses in the different morphological groups (Fishers exact test p=0.001).

The number of locules in the multilocular and multilocular solid categories ranged from two to greater than ten with a median of three (Table 4.6). There was no difference in the number of locules between the benign and malignant masses (Mann-Whitney U Test p=0.55).

Table 4.6: Locularity of adnexal masses

	Unilocular	Unilocular solid	Multi-locular	Multi-locular solid	Solid	Unclassifiable
Benign	43	12	41	19	19	3
Malignant	6	4	2	13	7	1
Total	**49**	**16**	**43**	**32**	**26**	**4**

The echogenicity of the cyst fluid was found to be anechoic in 37% of the masses and low level in 25% (Table 4.7). There was no significant difference in the echogenicity of cyst fluid in benign and malignant masses (Fishers Exact Test p=0.082)

Table 4.7: Echogenicity of cyst fluid

	Anechoic	Low level	Ground glass	Haemorrhagic	Mixed	Solid
Benign	54	30	15	2	29	7
Malignant	9	12	2	0	7	3
Total	**63**	**42**	**17**	**2**	**36**	**10**

p = 0.082 (Fishers exact test)

Acoustic shadows were present in 15% of masses. Papillary proliferations were seen in 20.3% of the cases whereas ascites was only present in 9.3% (Table 4.8).

Table 4.8: Presence of binary ultrasound variables in benign and malignant masses

Variable present	Benign	Malignant	Test of significance	p value
Ascites	6	10	Fishers Exact	<0.001
Bilateral	22	13	Chi-Square	0.006
Smooth internal wall	106	13	Chi-Square	<0.001
Papillary projection	22	13	Chi-Square	0.006
Acoustic shadows	16	2	Chi-Square	0.53

4.4.3 Doppler variables

93 of the masses had blood flow detectable on colour Doppler imaging (54.7%). 4 of the avascular masses were malignant (5.2%) whereas 29 of the masses with detectable flow (30.8%) were malignant ($p<0.001$, Chi-Square test). The calculated variables from the Doppler waveform are summarised below (Table 4.9). The subjective colour flow score given for the whole mass was 2 (denoting minimal flow) in 41 masses, 3 (denoting moderate flow) in 27 and 4 (increased flow) in 24 of the masses. (Table 4.10)

Table 4.9: Doppler variables in benign, borderline and invasive masses

	Benign	Borderline	Invasive	P*
PSV (SD)	9.1 (13.4)	19.7 (21.1)	30.7 (21.7)	<0.001
TAMXV (SD)	6.0 (9.7)	14.3 (15.9)	20.6 (14.3)	<0.001
PI (SD)	2.1 (1.1)	1.5 (1.2)	0.74 (0.38)	<0.001
RI (SD)	0.83 (0.23)	0.66 (0.29)	0.56 (0.23)	<0.001

*significance calculated using Kruskal Wallis U test

Table 4.10: Distribution of colour score

Colour Score	Definition	Benign	Malignant	Total
1	No flow	74 (94.9%)	4 (5.1%)	78
2	Minimal flow	35 (85.4%)	6 (14.6%)	41
3	Moderate flow	20 (74.1%)	7 (25.9%)	27
4	Increased flow	8 (33.3%)	16 (66.7%)	24

P<0.001 Chi-Square test

4.4.4 Diagnosis of borderline tumours

All demographic, ultrasound and Doppler variables were analysed for discrimination between benign, borderline and invasive tumours. All Doppler variables (colour score, PSV, TAMXV, RI and PI) were significantly different between the three histological groups ($p<0.001$). The most significant variables were the Doppler variables, age, ascites, smooth wall, tumour diameter and the presence of bilateral lesions (Table 4.11).

Table 4.11: Significant variables in the discrimination between benign, borderline and invasive masses

Variable	Benign	Borderline	Invasive	Statistical test	p value
Age (mean)	42.4	48.3	54.5	One way ANOVA	0.002
Most common presenting symptom	Pain	Pain	Distension	Fisher's exact	0.026
Menopausal (%)	28	55	59	Fisher's exact	0.005
Flow present (%)	46	63	100	Chi square	<0.001
Tumour diameter /mm (mean)	77.4	119.5	106.4	One way ANOVA	0.001
Volume /mm^3	270.7	620.6	600.1	Kruskall Wallis	0.003
Papillary proliferations (%)	16	27	45	Fisher's exact	0.006

Variable	Benign	Borderline	Invasive	Statistical test	p value
Bilateral (%)	16	27	50	Fisher's exact	0.002
Ascites (%)	4	27	32	Fisher's exact	<0.001
Smooth (%)	77	64	27	Fisher's exact	<0.001

4.4.5 Subjective assessment of mass

Thirty two masses were subjectively found to be invasive or of borderline malignancy on assessment of both gray scale and Doppler appearance of the mass. One hundred and twenty one masses were felt to be benign in appearance; the remaining seventeen masses were difficult to classify. An attempt was made to identify the histological type of the tumour based on its gray scale and vascular appearance. Ovarian cancer was the most common subjective diagnosis (29 masses); four masses were classified as borderline ovarian tumours. Serous and mucinous cystadenoma were also commonly diagnosed (18 masses each) as were dermoids (23 masses) and endometriomata (16 masses). Twenty masses were difficult to classify into a histological type (Figure 4.1).

For the prediction of malignancy, a sensitivity of 72.7 with a specificity of 81.8% was achieved.

Figure 4.1: Subjective classification of the masses into histological types.

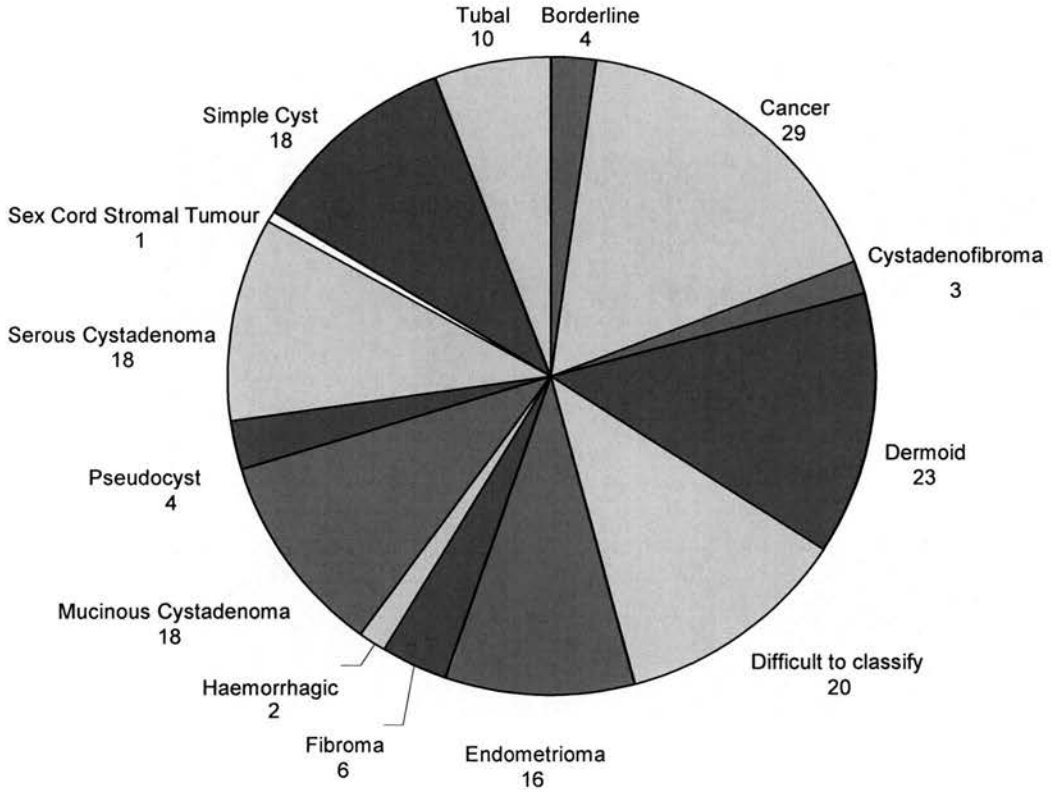


Table 4.12: Comparison of subjective impression to histology

Subjective Impression	Benign (%)	Borderline (%)	Invasive (%)	**Total**
Benign	112 (92.6)	7 (5.8)	2 (1.7)	**121**
Borderline	2 (50)	0	2 (50)	**4**
Invasive	9 (32.1)	2 (7.1)	17 (60.7)	**28**
Uncertain	**14 (82.4)**	**2 (11.8)**	**1 (5.9)**	**17**

$P < 0.001$ (Fishers exact test)

Table 4.13: Univariate analysis of all binary end points

Variable	Benign		Malignant		p
	Number	%	Number	%	
Nulliparity	36	26.3	10	30.3	0.457
Family history ovarian cancer	4	2.9	1	3.0	0.49
Family history breast cancer	6	4.4	1	3.0	0.46
Personal history breast cancer	6	3.6	2	6.1	0.38
Hysterectomy	11	8.0	6	18.0	0.155
Postmenopausal	38	27.7	19	57.6	0.001
Bilateral masses	22	16.1	13	39.4	<0.001
Ascites	6	4.4	10	30.3	<0.001
Papillary projection	22	16.1	13	39.4	0.006
Smooth internal wall	106	77.4	13	39.4	<0.001
Acoustic shadows	16	11.7	2	6.1	0.53

4.4.6 Prospective testing of models

Six models were tested prospectively: Jacobs' Risk of Malignancy Index; Tingulstad's Risk of Malignancy Index; Timmerman's logistic regression

model; Tailor's logistic regression model; Timmerman's neural network and the IOTA logistic regression model. The sensitivities ranged from 45.5 to 81.8 with specificities between 65.7 and 94.1 (Table 4.14). The performance of all models was poorer than in the original publications. Both risk of malignancy indices had a lower sensitivity for the detection of ovarian cancer in this data set than in the original study, however Jacobs' RMI performed better than Tingulstad's RMI and Timmerman's LRM. The neural network gave a good sensitivity and specificity but both were significantly lower than in the original publication. The IOTA model performed well with a good sensitivity and reasonable specificity. Subjective assessment of the mass gave a sensitivity of 72.7% and a specificity of 81.8% in the diagnosis of malignancy.

Table 4.14: Comparison of the prospective performance of different logistic regression models

Model	Publication date	Cut off	Accuracy in original study			Accuracy at validation			
			Sensitivity (%)	Specificity (%)	Area under ROC curve	Sensitivity (%)	Specificity (%)	Area under ROC curve	95% CI
Jacobs' RMI	1990	200	85	97	-	54.5	90.5	0.781	0.628, 0.865
Tingulstad RMI	1996	200	80	92	-	45.5	65.7	0.634	0.540, 0.729
Tailor LRM	1997	0.5	93	90	0.98	45.5	94.1	0.822	0.734, 0.910
Timmerman LRM	1999	0.25	94	83	0.97	54.5	78.8	0.728	0.697, 0.859
Timmerman ANN	1999	0.60	94	95	0.98	69.7	79.6	0.791	0.688, 0.893
IOTA LRM	2005	0.10	93	77	0.95	81.8	72.3	0.857	0.788, 0.926

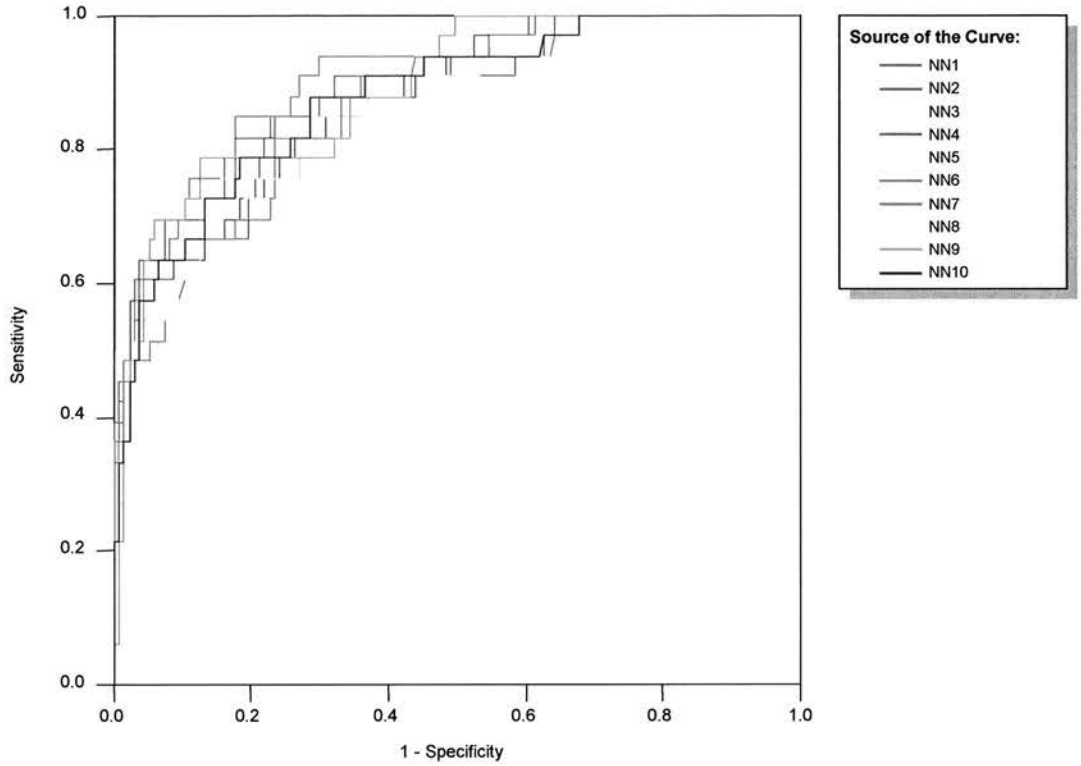
4.4.7 Artificial neural networks

The data set was randomly divided ten times into training and test sets. Multivariate logistic regression analysis of the ten different training sets yielded different significant variables. These are shown in Table 4.15. A neural network was developed from each of these sets of variables and applied to the each test set. The ROC curves for each neural network (NN) are shown in Figure 4.2.

Table 4.15: The significant variables in each data set

Data set	1st Variable	2nd Variable	3rd Variable	4th Variable	5th Variable
1	Menopause	Colour score	Bilateral	Internal wall	-
2	Volume	CA 125	Colour score	-	-
3	Tumour diameter	CA 125	Colour score	-	-
4	Tumour diameter	Colour score	-	-	-
5	Tumour diameter	Colour score	Bilateral	-	-
6	Menopause	Colour score	Bilateral	Ascites	CA 125
7	RI	Volume	CA 125	Ascites	-
8	PSV	Volume	Internal wall	Bilateral	Ascites
9	Age	Tumour diameter	Colour score	Bilateral	Ascites
10	Tumour diameter	Colour score	Bilateral	Ascites	-

Figure 4.2: The receiver operator characteristic curve for the ten morphological networks.



The best neural networks were NN1 and NN6. NN1 had four variable inputs menopausal status (meno), colour score (colour), bilaterality of masses (bilat) and presence of ascites (asc). There was one layer of three hidden nodes. Training of NN1 was halted after 910 iterations when the mean square error on the training and test sets were 0.375 and 0.282 respectively. The mathematical formula to calculate the prediction value of the network is:

$$\mathbf{Y = 1/(1+\exp(z_y))} \quad \mathbf{with} \quad \mathbf{z_y = 0.8403h_1 + 1.7496h_2 - 4.6023h_3 - 0.2953}$$

where:

$$h_1 = 1/(1+(\exp z_1)) \text{ with } z_1 = -1.4230(meno) + 1.1496(colour) + 0.8762(bilat) + 1.2193(asc) - 1.1187$$

$$h_2 = 1/(1+(\exp z_2)) \text{ with } z_2 = 0.1758(meno) + 0.8412(colour) - 0.4694(bilat) + 0.2071(asc) - 0.2436$$

$$h_3 = 1/(1+(\exp z_3)) \text{ with } z_3 = -2.062344(meno) - 0.8894(colour) - 0.7488(bilat) - 0.7515(asc) + 3.1671$$

With a probability of malignancy of 0.5, NN1 gave a sensitivity of 72.2% with a specificity of 90%. The area under the ROC curve was 0.894.

The sixth network (NN6) had five variable inputs menopausal status (meno), colour score (colour), CA 125/1000 (CA 125), bilaterality of masses (bilat) and presence of ascites (asc). There was one layer of three hidden nodes. Training was stopped after 3280 iterations when the test error and training error were 0.324 and 0.319 respectively.

The mathematical formula of the network is:

$$Y = 1/(1+\exp(z_y)) \text{ with } z_y = 4.8013h_1 + 4.3906h_2 - 4.0762h_3 + 0.4708$$

where:

$$h_1 = 1/(1+(\exp z_1)) \text{ with } z_1 = -2.6141(meno) + 3.5595(colour) - 0.6878(CA\ 125) + 2.4030(bilat) - 3.3074(asc) - 1.5492$$

$h_2 = 1/(1+(\exp z_2))$ with z_2 = -3.3334(meno) + 2.5809(colour) –1.4407(CA 125) + 1.5831(bilat) –3.37736(asc) –1.9517

$h_3 = 1/(1+(\exp z_3))$ with z_3 = -1.8254(meno) –0.8672(colour) –8.8538(CA 125) – 1.3481(bilat) +1.6700(asc) + 5.040331

At a probability of malignancy of 60%, the second neural network gave a sensitivity of 85.7% with a specificity of 88.6%. The area under the ROC curve was 0.904. The performance of the networks is summarised in Table 4.16.

Table 4.16: Comparison of the performance of all neural networks

Neural Network	Area under ROC curve	95% CI	
		Upper Bound	Lower Bound
1	0.894	0.833	0.955
2	0.877	0.811	0.942
3	0.872	0.803	0.942
4	0.872	0.803	0.942
5	0.850	0.783	0.918
6	0.904	0.847	0.960
7	0.865	0.797	0.933
8	0.865	0.795	0.934
9	0.887	0.826	0.948
10	0.878	0.811	0.944

There was no significant difference between the AUC for the models using the methods of Hanley and McNeill (1983).

4.5 Discussion

4.5.1 Univariate analysis

The variables with the best discrimination between benign and malignant masses were menopausal status, tumour volume, the presence of a smooth internal wall and ascites. 40.6% of multilocular solid tumours were malignant as compared to 4.7% of multilocular tumours without solid parts. This is a similar finding to Timmerman's study in 1999 and Valentin in 1994. All the Doppler variables including the colour score were significantly associated with malignancy. A larger tumour has an increased risk of malignancy and is more likely to present with pressure symptoms. This may explain the relation between malignancy and presentation with urinary symptoms or abdominal distension.

The presence of papillary proliferations and bilateral lesions were less significantly associated with malignancy in this dataset, compared to previous studies (Timmerman 2005, Tailor 1997). Echogenicity of cyst fluid and number of locules within the tumour were not predictive of malignancy in this study, in contrast to the IOTA study (Timmerman, 2005). The difference may be due to the considerably larger numbers in the IOTA study, enabling a true difference to be detected. A personal or family history of breast or ovarian cancer was unrelated to malignancy in this study. The numbers of women in these categories were small so a true difference may not have been identified.

The best variables at discriminating between benign, borderline and invasive masses were the Doppler variables, age, ascites, smooth wall, tumour diameter and the presence of bilateral lesions. These variables are similar to those with best discrimination between benign and malignant groups. Menopausal status, although significant, was less discriminatory than age in contrast to the discrimination between benign and malignant. The two variables are dependent but menopause is a binary endpoint whereas age is continuous. This is likely to explain why the difference between the ages of the groups is more significant than menopausal status.

4.5.2 Subjective assessment

Subjective assessment of the masses gave a sensitivity of 72.7% with a specificity of 81.8% in the diagnosis of malignancy. This is lower than that in other reported studies. Valentin produced a sensitivity of 88% with a specificity of 96%(Valentin, 1999). However she is an ultrasonographer of great experience. At the start of this study, the operator was a relatively inexperienced scanner, which may have contributed to the low sensitivity in this study. Timmerman's study in 1998 examining operator experience in subjective assessment demonstrated a higher accuracy with greater scanning experience. The operators in training achieved a sensitivity of 86.7 to 90.4% and specificity between 63.2 and 70.1%. These inexperienced scanners may have been advantaged by the prior selection of still images by the scanner performing the scan.

4.5.3 Prospective evaluation

This study has provided an external validation of six models. The IOTA LRM has not previously been validated; this study shows a strong prospective performance. Timmerman's neural network has previously been tested prospectively (Mol 2001), but considerable concerns about the study methodology were raised by the authors of the original study (Timmerman, 2001). This study has shown that diagnostic models perform less well when tested prospectively than in the studies in which they were developed. In the original publications, the sensitivities were all over 80% with specificities mostly over 90%. In this study, the sensitivities ranged from 45.5% to 82% with specificities of 66% to 94%.

Jacobs' RMI had a fair prospective performance with a lower sensitivity than in the original publication. A significant variation in the reported sensitivity exists in previous publications (Davies 1993, Tingulstad 1996, Morgante 1999, Timmerman 1999, Aslam 2000, Mol 2001, Timmerman 2005). The original paper did not describe a rigid classification for the ultrasound findings and therefore interpretation may vary between different centres and different ultrasonographers.

The three other models that have previously been tested prospectively gave a similar sensitivity and specificity to those found by other investigators (Morgante 1999, Aslam 2000, Valentin 2001). These results support the findings in this study.

Timmerman's neural network showed lower accuracy in discriminating between benign and malignant masses than in the original publication. Its performance was more accurate than that of the risk of malignancy indices.

This is an expected finding as the neural network provides a more complex non-linear fit to the data that cannot be achieved by the RMI or LRM models. It showed a higher sensitivity but similar specificity to both the Tailor LRM and the Timmerman LRM.

The IOTA LRM gave the best performance of all the prospectively tested models.

These findings can be explained by a number of factors. Models developed and tested on the same data set have an over-optimistic level of diagnostic accuracy. To make a model workable in clinical practice, only a small number of variables are used to make up the model. The selection of variables for insertion into a model is dependent on their discrimination between groups in the data set. Therefore, the development of the model is data-dependent and it provides a best fit for the examined data set. If the examined data set contains a small number of cases then the model is likely to perform poorly when applied to a new data set (Altman 2000). Timmerman's neural network was based on a data set of 173 patients. However the IOTA LRM was designed on a data set over 5 times larger. A model developed using a large number of examples gives a better generalisation and this partly explains the greater accuracy of the IOTA model.

Both Timmerman's LRM and ANN and the IOTA model all include colour score as a variable. This is a score which is awarded subjectively by the operator. The inclusion of a subjective variable decreases the reproducibility of the results obtained from these three models.

The clear definition of inclusion and exclusion criteria is important to define the sample data set. Excluded cases must be accounted for individually, as

cases excluded due to missing data may not be randomly spread through the data set (Altman 2000). Both Timmerman's ANN and the IOTA LRM detail the cases excluded from analysis. In development of the ANN, 18 patients were excluded (9.4%) because the level of CA 125 had not been measured. The IOTA data set had 83 exclusions (7%) due to missing data in 42 cases, mainly because no operation had been performed. Neither model appears to suffer from significant bias due to case exclusion.

The transportability of the models to a new centre with a different case-mix relates to their performance at prospective testing. The ANN was developed in one centre whereas the IOTA LRM was developed after a multicentre, international recruitment of patients. Because the IOTA model was designed on patients from different centres, it is not surprising that it performs well when tested prospectively on a new data set from a single centre.

In the original dataset of Timmerman's ANN, 28% of the masses were malignant. 26% of these were metastatic invasive tumours from other primary tumours (e.g. colon, endometrium or breast) and 10% were borderline tumours. In the IOTA data set, 25% of the cases were invasive. 15.8% were metastatic tumours and 20.7% were borderline. In the data set of this study, a similar proportion of cases were malignant. However all the invasive malignancies were primary and 36% were borderline tumours. The similarity of the IOTA data set to that of this study may also explain its good performance in this study. Both original papers described the methods of classification and examination technique in detail. Despite this, subtle differences are difficult to avoid and are likely to have affected the classification of papillary projection and colour score in this study. Although this would decrease the prospective accuracy of the models, it would affect

both models similarly and does not explain the difference in their performance.

The value of morphological models is questionable. This study has not confirmed the accuracy found in the original publications of the models. There is a large degree of intrinsic variability in this data set which may explain the variable performance of different models. A larger dataset would enable more robust conclusions to be drawn and stronger models to be developed. However, even the model designed on a large data set performed less well prospectively than in the original study. The IOTA model is complex and contains twelve different variables. Considerable sonographic expertise is needed to record these variables. Given the good performance of subjective assessment in the hands of an experienced sonographer, then this may be better than a statistical model.

4.5.4 Artificial neural networks

The data set was divided ten times to give ten different training and test sets. When each training set was analysed using logistic regression, different significant variables were produced for each training set. Overall, nine different variables were identified as significant (menopausal status, colour score, PSV, RI, tumour diameter, tumour volume, ascites, bilateral lesions and CA 125). The number of significant factors varied from two to five in different training sets. Ten neural networks were developed, one for each set of significant variables and tested on the respective test set. The performance of the networks was mixed with AUC's of between 0.85 to 0.904.

The finding of different significant variables in different randomly selected portions of the data set implies a large degree of heterogeneity in the cases. None of the variables was found to be significant in every training set. A wide variety of different histological types were included in the data set, which may explain the difficulty in classifying the cases. For example, the data set included twelve fibromas, a Brenner tumour, a benign steroid cell tumour and a Leydig cell tumour. These are all solid and richly vascularised and may be bilateral. Occasionally, they are accompanied by ascites with a modest rise in CA 125. They also all benign but these characteristics make correct classification difficult. Within the malignant masses, twelve of the thirty-three were borderline tumours. These are also notoriously difficult to classify (Aslam 2000, Valentin 2001).

This degree of variability within the data set suggests that the morphological variables are poor classifiers. In addition, the neural networks constructed using a selection of them are unlikely to generalise well when applied to a new population. To develop a good neural network, a larger data set would be required, given the large degree of variability within the case mix.

Subjective assessment performed as well as any of the models tested prospectively in this study, despite the relative lack of experience of the operator at the start of the study period. Given an experienced operator, none of the models offers a significant benefit. However, an inexperienced scanner may find the models useful to improve the diagnostic accuracy of his or her predictions.

Chapter 5

The application of models to incidentally found small masses

5.1 Background

Radiological imaging is increasingly becoming a first line investigation for women with a broad variety of abdomino-pelvic symptoms. Some gynaecologists consider that a transvaginal scan should be an integral part of every gynaecological examination (Valentin, 1999b). Women with pain or irregular vaginal bleeding are investigated with pelvic ultrasound. CT and MRI may be employed in cases of suspected renal calculi, persistent abdominal pain or back pain. The use of radiological investigations can lead to detection of adnexal masses that have not been clinically suspected. These masses may be found incidentally whilst scanning the abdomen for non-pelvic indications or during ovarian cancer screening. The women in whom these masses are detected are often asymptomatic and healthy. These findings will at best cause worry to the woman scanned. In many cases, additional further testing will be employed and at worst, they will lead to potentially harmful treatment. It is particularly important to be able to discriminate between malignant and benign masses in this population to avoid significant iatrogenic harm.

Bailey et al followed asymptomatic postmenopausal women with masses of less than 10 cm in maximum diameter. Two hundred and fifty six masses were unilocular, none of which was malignant at surgery or had developed cancer after 3 years follow-up. One hundred and fifteen masses were persistent and complex and 8% of these were malignant (Bailey, 1998).

Ferrazzi et al refined Lerner's morphological scoring system for the diagnosis of malignancy and applied it to a subset of masses with a mean diameter of 5 cm or less (Lerner 1994, Ferrazzi 1997). The new scoring system gave a

sensitivity of 92% and a specificity of 53% for the diagnosis of malignancy in the subset of small masses. This was tested prospectively in a multicentre study that gave a sensitivity of 92% and a specificity of 76.9% (Ferrazzi, 2005).

None of the published statistical models has been applied to small adnexal masses. All these models were developed on a heterogeneous group of masses, the bulk of which were large and clinically detectable (Jacobs 1990, Tingulstad 1996, Tailor 1997, Timmerman 1999). There are reasonable doubts as to whether these models can be successfully applied to small masses. The measurement of tumour diameter and volume are likely to be redundant in a population of small masses. The models tested in this study do not incorporate either tumour diameter or tumour volume as variables, so may be more likely to diagnose malignancy accurately than models that do encompass these variables.

5.2 Population

One hundred and seventy women were included in this study. Sixty-six women had masses smaller than 5 cm in mean diameter. None of these masses was detected clinically prior to the scan. The demographic, ultrasound and Doppler variables for this group were compared to the hundred and four women with a large, clinically detectable mass.

5.3 Results

5.3.1 Histology of small masses

Of the sixty-six patients with small masses, sixty were benign; three were borderline and three were invasive. The histology of the masses is shown in Table 5.1.

Table 5.1: Histology and stage of small adnexal masses

Histological Type	Nature	Stage	Number of masses
Endometrioma	Benign	-	10
Dermoid cyst	Benign	-	8
Serous cystadenoma	Benign	-	6
Mucinous cystadenoma	Benign	-	5
Fibroma	Benign	-	1
Simple ovarian cyst	Benign	-	14
Tubal	Benign	-	4
Cystadenofibroma	Benign	-	7
Miscellaneous	Benign	-	5
Serous cystadenoma	Borderline	I	2
Mucinous cystadenoma	Borderline	I	1
Serous cystadenocarcinoma	Invasive	III	1
Mucinous cystadenocarcinoma	Invasive	II	1
Leiomyosarcoma	Invasive	I	1
Total			66

5.3.2 Presenting symptoms

The most common indications for the scan were pelvic pain and bleeding. Eight masses were detected incidentally either at ovarian cancer screening or with another radiological modality e.g. CT or MRI. All the borderline and invasive masses presented with either pain or abnormal vaginal bleeding (Table 5.2).

Table 5.2: Presenting symptoms of small masses

	Benign	**Borderline**	**Invasive**	All histology
Incidental	8	0	0	8
Pain	32	2	1	35
Bleeding	19	1	2	22
Abdominal distension	1	0	0	1
Total	60	3	3	66

5.3.3 Univariate analysis

There was a difference in the presence of ascites, papillary proliferations, detectable flow and the smoothness of the internal wall between benign and malignant small cysts. Other variables were not significantly different between the groups (Table 5.3).

Table 5.3: Demographic and morphological variables by histology of mass

Variable	Benign n=60	Malignant n=6	Significance p
Mean age (Standard deviation)	41.9 (13.9)	46.8 (24.2)	0.641
Mean diameter/mm (Standard deviation)	42.0 (12.6)	44.5 (8.6)	0.195
Menopausal (%)	28	50	0.271
Detectable blood flow (%)	47	100	0.013
Simple unilocular cyst (%)	47	33.3	0.188
Papillary proliferation (%)	13.3	50	0.022
CA 125 < 35 U/ml (%)	16.7	33.3	0.199
Acoustic shadows (%)	10	16.7	0.613
Bilateral (%)	15	33.3	0.251
Ascites (%)	1.67	33.3	<0.001
Smooth (%)	75	33.3	0.032

5.3.4 Comparison of large and small masses

There were no significant differences between any of the demographic variables in women with large and small masses (Table 5.4). Significantly more of the small masses were unilocular and a greater proportion of the large masses were multilocular solid (Table 5.5). None of the small tumours presented with urinary symptoms and, unsurprisingly, significantly less presented with abdominal distension (Table 5.6).

Table 5.4: Demographic and morphological variables in large and small masses

Variable	Large masses n=104	Small masses n=66	Significance p
Age (mean and standard deviation)	45.6 (16.2)	42.3 (14.9)	0.175
Postmenopausal (%)	36	30	0.478
Hysterectomy (%)	12	8	0.401
Flow (%)	55	51	0.579
Acoustic shadows (%)	11	11	0.593
Papillary (%)	23	17	0.338
Bilateral (%)	23	17	0.338
Ascites (%)	13	5	0.108
Smooth (%)	69	71	0.864

Table 5.5: Presenting symptoms of large and small masses

Presenting Symptom	Large masses n=104	Small masses n=66
Bleeding (%)	22	35
Distension (%)	22	3
Incidental (%)	11	12
Pain (%)	38	50
Urinary (%)	7	0

Chi Square Test p=0.001

Table 5.6: Locularity of large and small masses

Locularity	Large masses n=104	Small masses n=66
Unilocular (%)	18	45
Unilocular Solid (%)	9	11
Multilocular (%)	30	18
Multilocular Solid (%)	25	9
Solid (%)	17	12
Unclassifiable (%)	1	5

Chi Square Test p=0.001

5.3.5 Subjective assessment

Subjective assessment gave a sensitivity of 83% in the diagnosis of malignancy. A correct diagnosis of the specific type of adnexal mass was made in 51% of cases. Dermoid cysts and endometriomas were more correctly classified than fibromas, tubal cysts and cystadenomas (Table 5.7).

Table 5.7: Accuracy of subjective assessment in the specific diagnosis of masses

Diagnosis	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Malignancy	83.3	78.3	27.7	97.9	80.3
Cystadenoma	50	89.6	64.3	82.7	78.8
Dermoid	100	98.3	88.9	100	98.5
Endometrioma	70	96.4	77.8	94.7	92.4
Sex cord tumour	50	100	100	98.5	98.5
Fibroma	33.3	98.4	50	96.9	95.5
Tubal	50	96.6	60	95.8	92.4

5.3.6 Prospective performance of statistical models

The performance of the statistical models in the small masses and in the whole dataset of 170 masses is shown in Table 5.8. Tingulstad's RMI and Timmerman's LRM and ANN were the best performing models. Jacobs' RMI at a cut off of 200 and Tailor's LRM had excellent specificity but only diagnosed one of the six malignancies. At a cut off of 100, Jacobs' model had an increased sensitivity to 33.3% with a slight sacrifice in specificity.

Table 5.8: Performance of models on small masses

Model	Publication Date	Cut Off	Variables in model	Small masses (n=66)		Whole dataset (n=170)	
				Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Jacobs RMI	1990	200	Ultrasound score, menopausal score, CA 125	16.7	98.3	54.5	90.5
Jacobs RMI	1990	100	Ultrasound score, menopausal score, CA 125	33.3	93.3	60.6	85.4
Tingulstad RMI	1996	200	Ultrasound score, menopausal score, CA 125	83.3	60	45.5	65.7
Tailor LRM	1997	0.5	Age, papillary projection, TAMXV	16.7	93.3	45.5	94.1
Timmerman LRM	1999	0.25	Menopausal score, papillary projection, CA 125, colour score	66.7	81.6	54.5	78.8
Timmerman ANN	1999	0.6	Menopausal status, CA 125, ascites, unilocularity, smoothness of internal wall, presence of papillary structures and bilateral masses	66.7	83.3	69.7	79.6
Subjective	-	-	-	83.3	78.3	72.7	81.8

5.4 Discussion

The incidence of malignancy was lower in the group of small masses (9.1%) than in the group of large masses (19.4%). This is consistent with a previous study that found a 7.7% incidence of malignancy in a similar population (Ferrazzi, 2005).

Two of the six malignant masses in this study were unilocular cysts. One was a stage IIa mucinous cystadenocarcinoma, the other a stage Ia borderline mucinous cystadenoma. They both demonstrated moderate vascularity with low level echogenicity and a CA 125 value within the normal range. This finding contrasts with previous studies showing an extremely low incidence of malignancy in unilocular cysts (Bailey 1998, Valentin 2002). This may be due to the smaller size of the masses in this study. This finding reinforces the need for follow-up in a small, incidentally diagnosed ovarian cyst.

The best discriminatory variables between benign and malignant small cysts were ascites, papillary proliferations, detectable flow and the smoothness of the internal wall. This contrasts with the univariate analysis on the whole dataset. Age, tumour diameter, menopausal status, bilateral lesions and tumour volume were all significant variables in large masses.

Some of the diagnostic models performed well in the small mass group. Both Jacobs' RMI and Tailor's LRM gave a lower sensitivity in the small compared to the large masses. This is likely to be due to a number of factors. The RMI is based on multiplication of the CA 125 level, menopausal score and ultrasound score, which varies from 0 to 3. Any mass scoring zero on the ultrasound criteria is therefore classed as benign, whatever the menopausal status and CA 125 level. Thus, a unilocular cyst with an

irregular internal wall, papillary proliferations and pronounced vascularity will be classified as benign. In addition, CA 125 was raised in only two of the six malignant masses and menopausal status was not a significant variable. This led to a low sensitivity but an impressive specificity of 98.3% in the small mass group. The sensitivity was improved to 33% by lowering the cut-off to 100 with only a small drop in specificity.

Tailor's LRM gave a good specificity but a low sensitivity in the small masses. Age was not significantly different between the benign and malignant masses. This is likely to be the main factor in the poor performance of the Tailor model as papillary projections and vascularity were significantly different between the groups.

Tingulstad's RMI has an ultrasound score ranging from 1 to 4 and thus does not suffer from a similar lack of sensitivity. The RMI performed better on the small mass group than on the whole dataset despite its retention of menopausal status and CA 125 as variables.

Both of Timmerman's models performed well in the small mass group. The LRM included the papillary projection score and colour score, both of which were significant variables. The ANN included three significant variables (ascites, smoothness of internal wall and the presence of papillary structures) but also four non-significant variables (menopausal status, CA 125, unilocularity and bilaterality). This gave the ANN a small improvement in specificity over the LRM but still detected only 4 out of 6 malignant masses.

All the models except Jacobs' RMI and Tailor's model performed at least as well in the small mass group as they did in the whole dataset. Subjective assessment was also as effective at discriminating between benign and malignant masses in the small mass group. This is a somewhat surprising

finding. Smaller masses may be expected to be more troublesome to characterise both subjectively and using models. This study has demonstrated that demographic variables such as age and menopausal status are non-discriminatory and CA 125 was not elevated in four of the six malignant masses. However, morphological and Doppler variables still allowed accurate classification of the majority of the masses.

Subjective assessment of the mass was better than the models in the diagnosis of ovarian cancer. The correct specific diagnosis was made in 51% of cases, which is consistent with previous publications (Valentin 2001, Jermy 2001). In this study, a sensitivity of 83.3% was achieved in the diagnosis of malignancy. This equates well to other studies published on the diagnosis of larger masses in which sensitivity ranges from 82 to 98% (Timmerman 1999, Valentin 2001). Subjective analysis has been shown to improve with increasing experience of the sonographer (Timmerman, 1999). Statistical models may help less experienced sonographers to improve their diagnostic accuracy but rigorous adherence to the described definitions is essential.

The hypothesis that the published models can be applied to small asymptomatic masses is supported by the findings for Timmerman's models and Tingulstad's RMI. Subjective diagnosis has been shown to be superior to all the models in the diagnosis of ovarian cancer in small asymptomatic masses.

Chapter 6

General Discussion

6.1 Introduction

This chapter summarises the results obtained in this thesis. Detailed discussion accompanied each chapter and therefore only the wider context of this work is discussed here. Finally, suggestions for further study are presented.

6.2 Synopsis of Results

The aim of this thesis has been to investigate the use of tumour markers and mathematical models in the preoperative diagnosis of ovarian cancer. Work has concentrated on the development of new mathematical models, the validation of previously published models, and the application of these models to incidentally diagnosed small masses.

The tumour markers CA 125, CA 15-3 and CA 72-4 were found to be valuable markers in the discrimination of benign from non-benign masses and were used to construct new mathematical models. Her-2/neu and CA 19-9 were not discriminatory for the diagnosis of ovarian malignancy. Two decision tree models were developed, one for the differentiation of benign from non-benign masses, the second for classifying benign, borderline and malignant masses. The retained variables in the two trees were CA 125 and CA 15-3. Both models achieved an excellent specificity for the diagnosis of malignancy. The sensitivity of the first model was 56.3% and the second 67%. As found in previous studies, borderline tumours were difficult to classify (Aslam 2000, Valentin 2001). Two neural networks were constructed using tumour markers and demographic variables and these were shown to give superior discrimination to the decision tree analysis. The first network used CA 125, CA 15-3 and age as inputs and obtained a sensitivity of 100% and a specificity of 90.9% on the test set. This outperformed the only previously published network to use tumour markers (Zhang 1999).

Two models were validated clinically in this study: the IOTA model and Timmerman's neural network. Neither of these models had been

satisfactorily validated in the literature. The best performing model at prospective validation was found to be the IOTA model. This had a higher sensitivity than all the other tested models and maintained a good specificity. The neural network also performed well with an adequate sensitivity and good specificity. In contrast, Jacobs' risk of malignancy index achieved a low sensitivity in this study. A significant variation in its performance has been found previously (Davies 1993, Tingulstad 1996, Morgante 1999, Timmerman 1999, Aslam 2000, Mol 2001, Timmerman 2005). Subjective assessment was as good as the best model in the discrimination of benign from non-benign masses.

Ten neural networks were developed using morphological parameters. The data set was randomly divided into training and test set ten times. For each division of the data, different morphological variables were significant on regression analysis. A network was developed for each set of variables. The performance of these networks was no better than either the validated models or subjective assessment.

The models were applied to a group of incidentally diagnosed small masses. Timmerman's neural network and logistic regression model classified the masses accurately whereas Jacobs' RMI and Tailor's LRM gave a low sensititivity in the diagnosis of malignancy. Subjective assessment was a better discriminator than any of the models in this group.

The value of morphological models is questionable. Their accuracy has not been confirmed in this thesis. Subjective assessment performed as well than any of the morphological models both in the large and small masses. The

new tumour marker models in this study were at least as good as the morphological models.

A comparison of all the models tested and developed in this thesis is displayed in Table 6.1.

Table 6.1: An overview of the performance of all models investigated in the thesis.

Model	Application	Variables	Sensitivity	Specificity	AUC	95% CI
Jacobs' RMI	Prospective	CA 125, Menopause, Ultrasound score	54.5	84.7	0.621	0.528 0.715
Tingulstad RMI	Prospective	CA 125, Menopause, Ultrasound score	45.5	65.7	0.634	0.540 0.729
Tailor LRM	Prospective	Age, TAMXV, Papillary projection	45.5	94.1	0.822	0.734 0.910
Timmerman LRM	Prospective	Menopause, CA 125, Papillary projection, Colour score	54.5	78.8	0.728	0.697 0.859
Timmerman NN	Prospective	Menopause, CA 125, Ascites, Unilocularity, Smooth wall, Papillary projection, Bilateral	69.7	79.6	0.791	0.688 0.893
IOTA LRM	Prospective	History of ovarian cancer, Hormonal use, Age, Tumour diameter, Pain, Ascites, Papillary projection with flow, Entirely solid tumour, Diameter of solid part, Irregular internal wall, Acoustic shadows, Colour score	81.8	72.3	0.857	0.78 0.926
Lawrence NN1	Retrospective	Menopause, Colour score, Bilateral, Ascites	72.2	90	0.894	0.833 0.955
Lawrence TNN1	Retrospective	Age, CA 125, CA 15-3	100	90.9	0.959	0.893 1.00

Model	Application	Variables	Sensitivity	Specificity	AUC	95% CI
Lawrence TNN2	Retrospective	Age, CA 125, CA 15-3, CA 72-4	60	90.9	0.85	0.71 0.99
Subjective assessment	-	-	72.7	81.8	-	-

6.3 Discussion: The wider context

Research is undertaken to further our knowledge and understanding of the environment. Within medical research it is hoped that the understanding gained may lead to benefit for patients. The studies in this thesis increase our understanding of an aspect of the diagnosis of ovarian cancer. As radiological investigations become more commonly and widely available, the correct diagnosis of incidentally found masses is of greater concern. The validation of models in the classification of these masses has opened up the possibility of confident preoperative diagnosis. This knowledge could be used to plan conservative or minimally invasive surgery. The potential risks of major abdominal surgery and the psychological sequelae of a potential cancer diagnosis could then be avoided.

The studies in this thesis have demonstrated that subjective assessment of a mass by an experienced sonographer is at least as good as statistical models in the diagnosis of malignancy. Expertise in ultrasound is therefore an important skill in gynaecology. Ultrasound training for gynaecologists enables dissemination of experience between individuals and centres. The Royal College of Obstetricians and Gynaecologists in the United Kingdom has recently developed a Special Skills Module in Gynaecological Ultrasound to promote training in this area. For those scanners who have not yet accrued significant experience, statistical models can aid diagnosis.

A limitation of this thesis is the number of women included in the studies. Although the numbers are greater than most published series, they are

nonetheless small. The results presented here must therefore be interpreted within this context.

The ability to detect ovarian cancer at an early stage is an attractive concept. A number of trials screening for ovarian cancer have investigated this potential (Jacobs 1999, Skates 2003). Due to the significant morbidity of surgery for presumed ovarian cancer, a screening programme must minimise its false positive diagnoses. The work in this thesis could be used to create a second-line test of expert ultrasonography. This could be offered to screen-positive women in order to exclude benign adnexal masses.

6.4 Suggestions for future studies

The tumour markers models developed in this thesis enabled accurate preoperative discrimination of benign from malignant masses. These models should be validated in a prospective, multicentre study. The analysis of large numbers of small proteins in the serum of patients with ovarian malignancy has become possible through the use of mass spectroscopy (Petricoin 2003). This emerging field of proteomics may yield valuable markers that could be incorporated into statistical models.

The use of tumour markers could be combined with ultrasound to increase their specificity. The use of subjective analysis could exclude those benign tumours with high levels of tumour markers such as endometriomas and dermoids. The subsequent application of tumour markers would yield a lower number of false positive results. Alternatively, subjective analysis could be used to characterise the tumours into specific types. The tumour marker models could then be applied to those masses found difficult to characterise.

Measurement of the learning curves of trainee ultrasonographers may enable more targeted training in the diagnosis of ovarian malignancy. A cut off experience level may be identified above which subjective assessment of the tumour is more accurate than application of published models.

New ultrasound modalities have been developed that may improve the sensitivity of imaging of blood flow in tumours. The use of three-dimensional colour Doppler assessment to visualise the spatial geometry of vessels may

improve the detection of malignancy. Intravenous contrast agents such as microbubbles may allow dynamic quantification of blood flow within a mass. Studies on hepatic imaging have shown increased sensitivity in the diagnosis of metastases and the detection of smaller lesions than previously possible (Blomley, 2001). A pilot study in ovarian cancer has shown promising results (D'Arcy, 2004).

Bibliography

Alagoz, T., Buller, R.E., Berman, M. 1994. "What is a normal CA125 level?" *Gynecologic Oncology* 53, 93-7.

Alcazar, J. and Jurado, M. 1999. Using a logistic regression model to predict malignancy of adnexal masses based in menopausal status, ultrasound morphology and color doppler findings" *Gynecologic Oncology,* 69, 146-150.

Alcazar, J., Errasti, T., Laparte, C., Jurado, M., and Lopez-Garcia, G. 2001. "Assessment of a new logistic model in the preoperative evaluation of adnexal masses" *Journal of Ultrasound in Medicine,* 20[8], 841-848.

Altman, D.G. 1998. "Practical statistics for medical research".

Altman, D.G. and Royston, P. 2000. "What do we mean by validating a prognostic model?" *Statistics in Medicine,* 19, 453-473.

Aslam, N., Ong, C., Woelfer, B., Nicolaides, K., and Jurkovic, D. 2000a. "Serum CA125 at 11-14 weeks of gestation in women with morphologically normal ovaries" *British Journal of Obstetrics and Gynaecology,* 107[5], 689-690.

Aslam, N., Tailor, A., Lawton, F., Carr, J., Savvas, M., and Jurkovic, D. 2000b. "Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer" *British Journal of Obstetrics and Gynaecology,* 107[11], 1347-1353.

Aslam, N., Banerjee, S., Carr, J., Savvas, M., Hooper, R., and Jurkovic, D. 2000c. "Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer" *Obstetrics and Gynecology,* 96[1], 75-80.

Axtell, A., Lee, M., Bristow, R. et al "Multi-instiutional reciprocal validation study of computed tomography predictors of suboptimal primary cytoreduction in patients with advanced ovarian cancer" *Journal of Clinical Oncology,* 28[4], 384-389.

Bailey, C., Ueland, F., Land, G., DePriest, P., Gallion, H., Kryscio, R., and van, Nagell, J. 1998. "The malignant potential of small cystic ovarian tumors in women over 50 years of age" *Gynecologic Oncology,* 69[1], 3-7.

Bast R, Feeney M, Lazarus H, Nadler L, Colvin R, Knapp R. 1981. "Reactivity of a monoclonal antibody with human ovarian carcinoma" *Journal of Clinical Investigation,* 68, 1131-1337.

Bast, R., Klug, T., St John, E. 1983. "A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer" *New England Journal of Medicine,* 309, 883-7.

Bast, R. C. Jr, Knauf, S., Epenetos, A., Dhokia, B., Daly, L., Tanner, M., Soper, J., Creasman, W., Gall, S., and Knapp, R. C. 1991. "Coordinate elevation of serum markers in ovarian cancer but not in benign disease" *Cancer,* 68[8], 1758-1763.

Berchuck, A., Kamel R, and Whitaker R. 1990. "Overexpression of HER-2/*neu* is associated with poor survival in advanced epithelial ovarian cancer" *Cancer Research,* 50, 4087-4091.

Berek, J. 2005. "Epithelial ovarian cancer." In: *Practical Gynecologic Oncology* (fourth edition) Lippincott, Wiliams and Wilkins.

Biagiotti, R., Desii, C., Vanzi, E., and Gacci, G. 1999. "Predicting ovarian malignancy: application of artificial neural networks to transvaginal and color Doppler flow US" *Radiology,* 210[2], 399-403.

Bjorkholm, E., Lundell, M., Gyftodimos, A, and Silfversward, C. 1990. "Dysgerminoma: the Radiumhemmet series 1927 - 1984" *Cancer,* 65, 38-44.

Blomley, M., Cooke, J., Unger, E., Monaghan, M., and Cosgrove, D. 2001. " Science, medicine, and the future: Microbubble contrast agents: a new era in ultrasound" *British Medical Journal,* 322, 1222-1225.

Bon, G. G., Kenemans, P., Verstraeten, R., van Kamp, G. J., and Hilgers, J. 1996. "Serum tumor marker immunoassays in gynecologic oncology: establishment of reference values" *American Journal of Obstetrics and Gynecology,* 174[1 Pt 1], 107-114.

Bourne, T.H., Campbell, S., Steer, C., Whitehead, M., and Collins, W.P. 1989. "Transvaginal colour flow imaging: a possible new screening technique for ovarian cancer" *British Medical Journal,* 299, 1367-1370.

Breiman, L., Friedman, J., Stone, C.J., Olshen R.A. 1984. Classification and Regression Trees. CRC Press. ISBN: 9780412048418

Campbell, S., Bhan, V., Royston, P., Whitehead, M., and Collins, W.P. 1989. "Transabdominal ultrasound screening for early ovarian cancer" *British Medical Journal,* 299, 1363-1367.

Canney, P., Moore, M., Wilkinson, P., James, R. "Ovarian cancer antigen CA125: a prospective clinical assessment of its role as a tumour marker" *British Journal of Cancer*, 50, 765-9. 1984.

Carter, W., Hoying, J., Boswell, C., and Williams, S. K. 2001. "HER-2/neu over-expression induces endothelial cell retraction" *International Journal of Cancer*, 91, 295-299.

Cheli, C. D., Morris, D. L., Kish, L., Goldblatt, J., Neaman, I., Allard, W. J., Yeung, K. K., Wu, A. H., Moore, R., Chan, D. W., Fritsche, H. A., Schwartz, M. K., and Very, D. L., Jr. 1998. "Multicenter evaluation of the Bayer Immuno I CA 15-3 assay" *Clinical Chemistry*, 44[4], 765-772.

Clayton, R. D., Snowden, S., Weston, M. J., Mogensen, O., Eastaugh, J., and Lane, G. 1999. "Neural networks in the diagnosis of malignant ovarian tumours" *British Journal of Obstetrics and Gynaecology*, 106[10], 1078-1082.

Corney, R., Everett, H., Howells, A., and Crowther, M. 1992. "The care of patients undergoing surgery for gynaecological cancer, the need for information, emotional support and counselling" *Journal of Advanced Nursing*, 17, 667-671.

D'Arcy, T., Jayaram, V., Lynch, M., Soutter, W.P., Cosgrove, D.O., Harvey, C.J., and Patel, N. 2004. "Ovarian cancer detected non-invasively by contrast-enhanced power Doppler ultrasound" *British Journal of Obstetrics and Gynaecology*, 111[6], 619-622.

Davies, A. P., Jacobs, I., Woolas, R., Fish, A., and Oram, D. 1993. "The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index" *British Journal of Obstetrics and Gynaecology*, 100[10], 927-931.

Dede, M., Gungor, S., Yenen, M., Alanbay, I., Duru, N., and Hasimi, A. 2006. "CA19-9 may have clinical significance in mature cystic teratomas of the ovary" *International Journal of Gynecological Cancer*, 16, 189-193.

DePriest P.D., van Nagell J.R., and Gallion H. 1993. "Ovarian cancer screening in asymptomatic postmenopausal women" *Gynecologic Oncology*, 51, 205-209.

DePriest, P.D., Gallion, H.H., Pavlik, E.J., Kryscio, R. J., and van Nagell, J.R. 1997. "Transvaginal sonography as a screening method for the detection of early ovarian cancer" *Gynecologic Oncology*, 65, 408-414.

Devine, P. L., McGuckin, M. A., Quin, R. J., and Ward, B. G. 1994. "Serum markers CASA and CA 15-3 in ovarian cancer: all MUC1 assays are not the same" *Tumour Biology*, 15[6], 337-344.

Dina, R., Rustin, G. J. S., Soutter, W. P. 2003. "Carcinoma of the ovary and fallopian tube" In: *Shaw, R.W., Soutter, W. P., Stanton, S.L., eds. Gynaecology*, (third edition), Churchill Livingstone.

Doppler, C. "Ueber das farbige licht der doppelsterne und einiger anderer gestime des himmels" Abhandl.d.konigl. *Bohmischen Gesellschaft der Wissenschaften Sex* 2, 465-482. 1843 (translation).

Donald, I., MacVicar, J., Brown, T.G. 1958. "Investigation of abdominal masses by pulsed ultrasound" *Lancet* 1, 1188-1195.

Drapkin, R., Crum, C. P., and Hecht, J. L. 2004. "Expression of candidate tumor markers in ovarian carcinoma and benign ovary: evidence for a link between epithelial phenotype and neoplasia" *Human Pathology*, 35[8], 1014-1021.

Duffy, M.J. 1998. "CA 19-9 as a marker for gastrointestinal cancers: a review" *Annals of Clinical Biochemistry*, 35, 364-370.

Einhorn, N. and Zurawski, V. 1989. "CA 125 assay used in conjunction with CA 15-3 and TAG-72 assays for discrimination between malignant and non-malignant diseases of the ovary" *Acta Oncology*, 28[5], 655-657.

Einhorn, N., Sjovall, K., and Knapp, R. C. 1986. "Prospective evaluation of serum CA125 for early detection of ovarian cancer" *Obstetrics and Gynecology*, 80, 14-18.

Eltabbakh, G.H., Belinson, J.L., Kennedy, A.W., Biscotti, C., Casey, G., and Tubbs, R.R. 1996. "p53 and HER-2/*neu* overexpression in ovarian borderline tumours" *Gynecologic Oncology*, 65, 218-224.

Engelen, M.J., de Bruijn, H.W., Hollema, H., ten Hoor, K.A., Willemse, H., Aalders, J.G., and vander Zee, A.G. 2000. "Serum CA 125, carcinoembryonic antigen, and CA 19-9 as tumor markers in borderline ovarian tumors" *Gynecologic Oncology*, 78, 16-20.

Feng, H., Ghazizadeh, M., Konishi, H., and Araki, T. 2002. "Expression of MUC1 and MUC2 mucin gene products in human ovarian carcinomas" *Japanese Journal of Clinical Oncology*, 32[12], 525-529.

Ferrazzi, E., Lissoni A, and Dordoni D. 2005. "Differentiation of small adnexal masses based on morphologic characteristics of transvaginal sonographic imaging" *Journal of Ultrasound in Medicine,* 24, 1467-1473.

Ferrazzi, E., Zanetta G, Dordoni D, Berlanda N, Mezzopane R, and Lissoni G. 1997. "Transvaginal ultrasonographic characterisation of ovarian masses: comparison of five scoring systems in a multicentre study" *Ultrasound in Obstetrics and Gynecology,* 10, 192-197

Fisher, R. A. 1936. "The use of multiple measurements in taxonomic problems" *Annals of Eugenics,* 7, 179-188

Flamm, F., Einhorn, N., and Sjovall, K. 1988. "Symptomatology of ovarian cancer" *European Journal of Obstetrics, Gynecology, and Reproductive Biology,* 27, 53-57.

Fleischer, A. C., Cullinan, J. A., Kepple, D. M., and Williams, L. L. 1993a. "Conventional and color Doppler transvaginal sonography of pelvic masses: a comparison of relative histologic specificities" *Journal of Ultrasound in Medicine,* 12[12], 705-712.

Fleischer, A. C., Rodgers, W. H., Kepple, D. M., Williams, L. L., Jones, H. W. 3rd, FA Fleischer, A. C., and Jones, H. W. 1993b. "Color Doppler sonography of ovarian masses: a multiparameter analysis" *Journal of Ultrasound in Medicine,* 12[1], 41-48.

Folkman, J. "Fundemental concepts of the angiogenic process" *Current Molecular Medicine,* 3, 643-651. 2003.

Gadducci, A., Ferdeghini, M., and Prontera, C. 1992. "The concomitant determination of different tumor markers in patients with epithelial ovarian cancer and benign ovarian masses: relevance for differential diagnosis" *Gynecologic Oncology,* 44, 147-154.

Gero, E.J. 1989. "The CA72-4 radioimmunoassay for the detection of the TAG-72 carcinoma associated antigen in the sera of patients" *Journal of Clinical Laboratory Analysis,* 3, 360-369.

Gosling, R. and King, D. 1975. "Ultrasonic angiology" *In: Harens A, Adamson L, eds. Arteries and Vein,* 61-98.

Granberg, S., Wikland, M., and Jansson, I. 1989. "Macroscopic characterization of ovarian tumors and the relation to the histological diagnosis: criteria to be used for ultrasound evaluation" *Gynecologic Oncology,* 35[2], 139-144.

Granberg, S., Norstrom, A., and Wikland, M. 1990. "Tumors in the lower pelvis as imaged by vaginal sonography" *Gynecologic Oncology, 37*[2], 224-229.

Guadagni, F. and Roselli, M., Amato, T. 1992. "TAG-72 as a complemetary serum tumour antigen to CEA in monitoring patients with colorectal cancer" *Cancer, 72,* 2098-2106.

Guadagni, F., Marth, C., Zeimet, A. G., Ferroni, P., Spila, A., Abbolito, R., Roselli, M., Greiner, J. W., and Schlom, J. 1994. "Evaluation of tumor-associated glycoprotein-72 and CA 125 serum markers in patients with gynecologic diseases" *American Journal of Obstetrics and Gynecology,* 171[5], 1183-1191.

Haaften-Day, C., Russel, P., Boyer, C.M., Kerns, B.J., Wiener, J.R., Jensen, D.N., Bast, R.C., and Hacker, N.F. 1996. "Expression of cell regulatory proteins in ovarian borderline tumors" *Cancer, 77,* 2092-2098.

Hanley, J. and McNeil, B. 1983. "A method for comparing the areas under receiver operator characteristic curves derived from the same cases" *Radiology,* 148, 839-843.

Harrell, E., Lee, K.L., Calil, R.M., Pryor, D.B., and Rosati, R.A. Regression modelling strategies for improved prognostic modelling" *Statistics in Medicine,* 1984, 143-152.

Hasholzner, U., Baumgartner, L., and Stieber, P. 1996. "Clinical significance of the tumour markers CA125 II and CA72-4 in ovarian carcinoma" *International Journal of Cancer,* 69, 329-334.

Hata, K., Yoshida, M., Maruyama, R., Fujiwaki, R., and Miyazaki, K. 2002. " Prognostic significance of ultrasound derived intratumoural peak systolic velocity in epithelial ovarian cancer" *Ultrasound in Obstetrics and Gynecology,* 20, 186-191.

Itzkowitz, S.H., Yuan, M., Palekar, A., and Kim, Y. S. 1986. "Cancer-associated alterations of blood group antigen expression in human colorectal polyps" *Cancer Research,* 46[11], 5976-5984.

Jacobs, I., Stabile, I., Bridges, J., Kemsley, P., Reynolds, C., and Grudzinkas, J. 1999. "Multimodal approach to screening for ovarian cancer" *Lancet*, 268-271.

Jacobs, I., Oram, D., Fairbanks, J., Turner, J., Frost, C., and Grudzinskas, J. G. 1990. "A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer" *British Journal of Obstetrics & Gynaecology*, 97[10], 922-929.

Jacobs, I. J., Rivera, H., Oram, D. H., and Bast, R. C., Jr. 1993. "Differential diagnosis of ovarian cancer with tumour markers CA 125, CA 15-3 and TAG 72.3" *British Journal of Obstetrics and Gynaecology*, 100[12], 1120-1124.

Jauniaux E., Jurkovic D., Delogne-Desnoek J. and Meuris, S 1992. "Influence of human chorionic gonadotrophin, oestradiol and progesterone on uteroplacental and corpus luteum blood flow in normal early pregnancy" *Human Reproduction*, 7[10], 1467-1473.

Jermy, K., Luise, C., and Bourne, T. 2001. "The characterisation of common ovarian cysts in premenopausal women" *Ultrasound in Obstetrics & Gynecology*, 17, 140-144.

Junor, E.L., Hole, D.J., and Gillis, C.R. 1987. "Management of ovarian cancer: referral to a multidisciplinary team matters" *British Journal of Cancer*, 70, 363-370.

Kabawat, S., Bast, R., Bhan, A., Welch, W., Knapp, R., Colvin, R. 1983. "Tissue distribution of a coelomic-epithelium-related antigen recognised by the monoclonal antibody OC125" *International Journal of Pathology*, 2, 275-285.

Kaneko, Z., Kotani, H., Komuta, K., Satomura, S. 1961. "Studies on peripheral circulation by ultrasonic blood-rheograph" *Japanese Circulation Journal*, 25, 203-213.

Kehoe, S., Powell, J., Wilson, S., and Woodman, C. 1994. "The influence of the operating surgeon's specialisation on patient survival in ovarian carcinoma" *British Journal of Cancer*, 70[5], 1014-1017.

Kenemans, P., van Kamp, D., Oehr, P., Venstraeten, R. 1983. "Heterologous double-determinant immunoradiometric assay CA125 II: Reliable second generation immunoassay for determining CA125 in serum" *Clinical Chemistry*, 39(12), 2509 –2513.

Kobayashi, M. 1976. "Use of diagnostic ultrasound in trophoblastic neoplasms and ovarian tumors" *Cancer*, 38[1 SUPPL], 441-452.

Kudoh, K., Kikuchi, Y., and Kita, T. 1999. "Preoperative determination of several serum tumor markers in patients with primary epithelial ovarian carcinoma" *Gynecologic and Obstetric Investigation,* 47, 52-57.

Kurjak, A., Zalud, I., and Alfirevic, Z. 1991. "Evaluation of adnexal masses with transvaginal color ultrasound" *Journal of Ultrasound in Medicine,* 10[6], 295-297.

Kurjak, A., Shalan H, and Kupesic, S. 1994. "An attempt to screen asymptomatic women for ovarian and endometrial cancer with transvaginal color and pulsed Doppler sonography" *Journal of Ultrasound in Medicine,* 295-301.

Lerner, J. P., I.E., Federman, A., and Abramovich, G. 1994. "Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system" *American Journal of Obstetrics & Gynecology,* 170[1], 81-85.

Macdonald, F., Bird, R., Stokes, H., Russell, B., and Crocker, J. 1988. "Expression of CEA, CA125, CA19-9 and human milk fat globule membrane antigen in ovarian tumours" *Journal of Clinical Pathology,* 41[3], 260-264.

Maestranzi, S., Przemioslo, R., Mitchell, H., and Sherwood, R.A. 1998. " The effect of benign and malignant liver disease on the tumour markers CA19-9 and CEA" *Annals of Clinical Biochemistry,* 35, 99-103.

McKenzie, S., DeSombre, K., and Bast, B. 1993. "Serum levels of HER-2/*neu* correlate with overexpression of p185neu in human ovarian cancer" *Cancer,* 71, 3942-3946.

Meden, H., Marx, D., and Fattahi, A. 1994. "Elevated serum levels of a c-erbB-2 oncogene product in ovarian cancer patients and in pregnancy" *Journal of Cancer Research and Clinical Oncology,* 120, 378-381.

Meden, H., Marx, D., and Schauer, A. 1997. "Prognostic significance of p105 (c-erbB-2, HER2/*neu*) serum levels in patients with ovarian cancer" *Anticancer Research,* 17, 757-760.

Meire, H. B., Farrant, P., and Guha, T. 1978. "Distinction of benign from malignant ovarian cysts by ultrasound" *British Journal of Obstetrics and Gynaecology,* 85[12], 893-899.

Mol, B., Boll, D., De Kanter, M., Heintz, A., Sijmons, E., and Oei, S. 2007. "Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models." *Gynecologic Oncology,* 80, 162-167.

Morgante, G., Marca, A., Ditto, A., and De Leo, V. 1999. "Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses" *British Journal of Obstetrics and Gynaecology,* 106, 524-527.

Muramatsu, T., Mukai, M., and Sato, S. 2005. "Clinical usefulness of serum and immunohistochemical markers in patients with stage Ia and Ic ovarian cancer" *Oncology Reports,* 14, 861-865.

Nap, M. "Immunohistochemistry of CA125: unusual expression in normal tissues, distribution in the human fetus and questions around its application in diagnostic pathology" *International Journal of Biological Markers,* 13, 210-215. 1998.

NHS Executive. "A policy framework for commissioning cancer services" EL(95)51, Department of Health, 1995.

NHS Executive." Improving outcomes in gynaecological cancers" 16149 IP 14,550 Department of Health, 1999.

Ozaksit, G., Caglar, T., Cicek, N., Kuscu, E., Batioglu, S., and Gokmen, O. 1995. "Serum CA 125 levels before, during and after treatment for endometriosis" *International Journal of Gynaecology and Obstetrics,* 50[3], 269-273.

Parkes, C., Smith, D., Wald, N., and Bourne, T. 1995. "Feasibility study of a randomised controlled trial of ovarian cancer screening" *Journal of Medical Screening,* 1, 209-214.

Parkin, D., Whelan, S., and Ferlay, J. 1997. "Cancer Incidence in Five Continents, Volume VII" *IARC Scientific.*

Payne, R., Allard, J., Anderson-Mauser, L., Humphreys, J., Tenney, D., and Morris, D. 2000. "Automated assay for HER-2/neu in serum" *Clinical Chemistry,* 46, 175-182.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R., and Feinstein, A.R. 1996. " A simulation study of the number of events per variable in logistic regression analysis" *Journal of Clinical Epidemiology,* 49, 1373-1379.

Petricoin, E., Ardekani, A., and Hitt, B. 2003. "Use of proteomic patterns to identify ovarian cancer" *Lancet*, 359, 572-577.

Pourcelot, L. 1974. "Applications clinique de l'examen Doppler transcutane" *In: Peronneau P ed. Velocimetric ultrasonore Doppler.* 213-240.

Quinn, M., Babb, P., Brock, A., Kirby, I., and Jones, J. 2001. "Cancer trends in England and Wales 1950 to 1999" *Studies on medical and population subjects no.66, Stationery Office, London.*

Rosen, D.G., Wang, L., Atkinson, J.N., Yu, Y., Lu. K.H., Diamandis, E.P., Hellstrom, I., Mok, S.C., Liu, J., and Bast, R.C. Jr. 2005. "Potential markers that complement expression of CA125 in epithelial ovarian cancer" *Gynecologic Oncology*, 99, 267-277.

Safi, F., Kohler, I., Rottinger, E., and Beger, H. 1991. "The value of the tumor marker CA 15-3 in diagnosing and monitoring breast cancer" *Cancer*, 68, 574-582.

Sassone, A. M., I.E., Artner, A., Westhoff, C., and Warren, W. B. 1991. "Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy" *Obstetrics & Gynecology*, 78[1], 70-76.

Sato, S., Kobayashi, Y., and Okuma, Y.. 2002. " Establishment and characterization of a cell-line originated from human mucinous cystadenocarcinoma of the ovary" *Human Cell*, 15, 171-177.

Scambia, G., Benedetti Panici, P., Perrone, L., Sonsini, C., Giannelli, S., Gallo, A., and Mancuso, S. 1990. "Serum levels of tumour associated glycoprotein (TAG 72) in patients with gynaecological malignancies" *Cancer Letters*, 77, 145-153.

Schutter, E. M., Sohn, C., and Kristen, P. 1998. "Estimation of probability of malignancy using a logistic model combining physical examination, ultrasound, serum CA125 and serum CA72-4 in postmenopausal women with a pelvic mass: an international multicenter study" *Gynecologic Oncology*, 69, 56-63.

Skates, S. J., Menon, U., and MacDonald N. 2003. "Calculation of the risk of ovarian cancer from serial CA125 values for preclinical detection in postmenopausal women" *Journal of Clinical Oncology*, 21, 206-210.

Skates, S. J., Menon, U., and MacDonald N. 2003. "Calculation of the risk of ovarian cancer form serial CA125 values for preclinical detection in postmenopausal women" *Journal of Clinical Oncology*, 2, 206-210.

Skates, S. J., Horick, N., Yu, Y., Xu, F. J., Berchuck, A., Havrilesky, L. J., de Bruijn, H. W., van der Zee, A. G., Woolas, R. P., Jacobs, I. J., Zhang, Z., and Bast, R. C. Jr. 2004. "Preoperative sensitivity and specificity for early-stage ovarian cancer when combining cancer antigen CA-125II, CA 15-3, CA 72-4, and macrophage colony-stimulating factor using mixtures of multivariate normal distributions" *Journal of Clinical Oncology*, 22[20], 4037-4039.

Sladkevicius, P. and Valentin, L. 1995. "Interobserver agreement in the results of Doppler examinations of extrauterine pelvic tumors" *Ultrasound in Obstetrics and Gynecology*, 6[2], 91-96.

Sladkevicius, P., Valentin, L., and Marsal, K. 1995. "Transvaginal Doppler examination for the differential diagnosis of solid pelvic tumors" *Journal of Ultrasound in Medicine*, 14[5], 377-380.

Slamon, D. J., Godolphin, W., and Jones, L. A. 1989. "Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer" *Science*, 244, 707-712.

Stein, J. F. 1982. "An introduction to neurophysiology" *Oxford: Blackwell Science*, 346.

Tailor A, Jurkovic D, Bourne TH, Natucci M, Collins WP, and Campbell S. 1996. "A comparison of intratumoural indices of blood flow velocity and impedance for the diagnosis of ovarian cancer" *Ultrasound in Medicine and Biology*, 22, 837-843.

Tailor, A., Jurkovic, D., Bourne, T. H., Collins, W. P., and Campbell, S. 1997. "Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis" *Ultrasound in Obstetrics and Gynecology*, 10, 41-47.

Tailor, A., Jurkovic, D., Bourne, T. H., Natucchi, M, Collins, W. P., and Campbell, S. 1998. "Comparison of transvaginal color doppler imaging and color doppler energy for assessment of intraovarian blood flow" *Obstetrics and Gynaecology*, 91, 561-567

Tailor, A., Jurkovic, D., Bourne, T. H., Collins, W. P., and Campbell, S. 1999. "Sonographic prediction of malignancy in adnexal masses using an artificial neural network" *British Journal of Obstetrics and Gynaecology*, 106[1], 21-30.

Thor A. 1986. "Tumor-associated glycoprotein (TAG-72) in ovarian carcinomas defined by monoclonal antibody B72.3" *Journal of the National Cancer Institute,* 76, 995-1006.

Timmerman, D., Bourne, T. H., Tailor, A., Collins, W. P., Verrelst, H., Vandenberghe, K., and Vergote, I. 1999a. "A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model" *American Journal of Obstetrics and Gynecology,* 181[1], 57-65.

Timmerman, D., Verrelst, H., Bourne, T. H., De Moor, B., Collins, W. P., Vergote, I., and Vandewalle, J. 1999b. "Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses" *Ultrasound in Obstetrics and Gynecology,* 13[1], 17-25.

Timmerman, D., Schwarzler, P., Collins, W. P., Claerhout, F., Coenen, M., Amant, F., Vergote, I., and Bourne, T. H. 1999. "Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience" *Ultrasound in Obstetrics and Gynecology,* 13[1], 11-16.

Timmerman, D., Valentin, L., Bourne, T. H., Collins, W. P., Verrelst, H., Vergote, I., and International Ovarian Tumor Analysis, IOTA. 2000. "Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group" *Ultrasound in Obstetrics and Gynecology,* 16[5], 500-505.

Timmerman, D., Verrelst, H., Collins, W., Bourne, T., and Vergote, I. 2001. "Comment on: Distinguishing the benign and malignant mass: an external validation of prognostic models" *Gynecologic Oncology,* 166-167.

Timmerman, D. 2004. "The use of mathematical models to evaluate pelvic masses; can they beat an expert operator?" *Best Practice and Research in Clinical Obstetrics and Gynaecology,* 18, 91-104.

Timmerman, D., Testa, A., Bourne, T., Ferrazzi, E., Ameye, L., Konstantinovic, M., Van Calster, B., Collins, W., Vergote, I., van Huffel, S., and Valentin, L. 2005. "Logistic regression model to distinguish between the benign and malignant mass before surgery: a multicentre study by the international ovarian tumor analysis group" *Journal of Clinical Oncology,* 23, 8794-8801.

Tingulstad, S., Hagen, B., Skjeldestad, F. E., Onsrud, M., Kiserud, T., Halvorsen, T., and Nustad, K. 1996. "Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses" *British Journal of Obstetrics and Gynaecology,* 103[8], 826-831.

Udagawa, Y., Aoki, D., Ito K, Uejima T, Uemura M, and Nozawa, S. 2007. "Clinical characteristics of a newly developed ovarian tumour marker, galactosyltransferase associated with tumour (GAT)" *European Journal of Cancer,* 34, 489-495.

Valentin, L., Sladkevicius, P., and Marsal, K. 1994. "Limited contribution of Doppler velocimetry to the differential diagnosis of extrauterine pelvic tumors" *Obstetrics and Gynecology,* 83[3], 425-433.

Valentin, L. 1999a. "Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound" *Ultrasound in Obstetrics and Gynecology.,* 14[5], 338-347.

Valentin, L. 1999b. "High-quality gynecological ultrasound can be highly beneficial, but poor-quality gynecological ultrasound can do harm" *Ultrasound in Obstetrics and Gynecology,* 13[1], 1-7.

Valentin, L. 2000. "Comparison of Lerner score, Doppler ultrasound examination, and their combination for discrimination between benign and malignant adnexal masses" *Ultrasound in Obstetrics and Gynecology,* 15[2], 143-147.

Valentin, L., Hagen, B., Tingulstad, S., and Eik-Nes, S. 2001. "Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation" *Ultrasound in Obstetrics and Gynecology,* 18[4], 357-365.

Valentin, L. and Akrawi, D. 2002. "The natural history of adnexal cysts incidentally detected at transvaginal ultrasound examination in postmenopausal women" *Ultrasound in Obstetrics & Gynecology,* 20[2], 174-180.

Valentin, L. 2004. "Use of morphology to characterize and manage common adnexal masses" *Best Practice and Research in Clinical Obstetrics and Gynaecology,* 18[1], 71-89.

Valentin, L., Ameye, L., Testa, A., Lecuru, F., Bernard, JP., Paladini, D., Van Huffel, S., and immerman, D. 2006. "Ultrasound characteristics of different types of adnexal malignancies" *Gynecologic Oncology,* 102[1], 41-48.

van Nagell J.R., Higgins R, and Gallion H. 1990. "Transvaginal ultrasonography as a screening method for ovarian cancer. A report of the first 1000 cases screened" *Cancer,* 65, 573-577.

van Nagell J.R., DePriest P.D., and Puls LE. 1991. "Ovarian cancer screening in asymptomatic postmenopausal women by transvaginal sonography" *Cancer,* 68, 458-462.

Vaupel, P.W. "Blood flow, oxygenation, tissue pH distribution and bioenergetic state of tumours" *Ernst Schering Research Foundation,* No.23, Berlin, Germany.

Vestergaard, E.M., Hein H.O., Meyer, H., Grunnet, N., Wolf, H., Orntoft, T.F. 1999. "Reference values and biological variation for tumor marker CA 19-9 in serum for different Lewis and secretor genotypes and evaluation of secretor and Lewis genotyping in a Caucasian population" *Clinical Chemistry,* 45[1], 54-61.

Vergote, I., De Brabanter, J., Fyles, A., Bertelsen, K., Einhorn, N., Sevelda, P., Gore, M. E., Kaern, J., Verrelst, H., Sjovall, K., Timmerman, D., Vandewalle, J., Van Gramberen, M., and Trope, C. G. 2001. "Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma" *Lancet,* 357[9251], 176-182.

Walach, N. and Gur, Y. 1998. "Leukocyte alkaline phosphatase, CA15-3, CA125, and CEA in cancer patients" *Tumori,* 84[3], 360-363.

Weiss, N. and Peterson, A. 1978. "Racial variation in the incidence of ovarian cancer in the United States" *American Journal of Epidemiology,* 107, 91-95.

WHO Regional Office for Europe (2004). Highlights on health, United Kingdom, 2004. Copenhagen, WHO Regional Office for Europe

Williams, S., Blessing, J., Liao, S., Ball, H., and Hanjani, P. 1994. "Adjuvant therapy of ovarian germ cell tumours with cisplatin, etoposide and bleomycin: a trial of the Gynecological Oncology Group" *Journal of Clinical Oncology,* 12, 701-706.

Woolas R, Conaway M, Xu F, and Jacobs I. 1995. "Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses" *Gynecologic Oncology*, 59, 111-116.

Zhang, Z., Barnhill S, and Zhang, H. 1999. "Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses" *Gynecologic Oncology*, 73[1], 56-61.

Zurawski, V., Orjaseter, H., Andersen, A., and Jellum, E. 1988. "Elevated serum CA125 levels prior to diagnosis of ovarian neoplasia: relevance for detection of early ovarian cancer" *International Journal of Cancer*, 42, 677-680.

Appendix I: International Federation of Gynaecology and Obstetrics (FIGO) staging for primary carcinoma of the ovary

Stage	Stage	Description
Stage I		Growth limited to the ovaries
	IA	Growth limited to one ovary; no ascites containing malignant cells. No tumour on the external surface; capsule intact
	IB	Growth limited to both ovaries; no ascites containing malignant cells. No tumour on the external surface; capsule intact
	IC	Tumour either stage 1A or 1B but with tumour on the surface of one or both ovaries; or with capsule ruptured; or with ascites present containing malignant cells or with positive peritoneal washings
Stage II		Growth involving one or both ovaries with pelvic extension
	IIA	Extension and/or metastases to the uterus and/or tubes
	IIB	Extension to other pelvic structures
	IIC	Tumour either Stage IIA or IIB but with tumour on the surface of one or both ovaries; or with capsule ruptured; or with ascites present containing malignant cells or with positive peritoneal washings
Stage III		Tumour involving one or both ovaries with peritoneal implants outside the pelvis and/or positive retroperitoneal or inguinal lymph nodes. Superficial liver metastasis equals Stage III. Tumour is limited to the true pelvis but with histologically proven malignant extension to small bowel or omentum.
	IIIA	Tumour grossly limited to the true pelvis with negative nodes but with histologically confirmed microscopic seeding of abdominoperitoneal surfaces.
	IIIB	Tumour of one or both ovaries with histologically confirmed implants of abdominoperitoneal surfaces, none exceeding 2cm in diameter. Nodes negative.
	IIIC	Abdominal implants >2cm in diameter and/or positive retroperitoneal or inguinal nodes.
Stage IV		Growth involving one or both ovaries with distant metastasis. If pleural effusion is present, there must be positive cytologic test results to allot a case to Stage IV. Parenchymal liver metastasis equals Stage IV.

Appendix II: Information sheet on the value of biochemical markers and ultrasound in the management of ovarian cysts.

We are asking for your help with research that is being carried out in our hospital. We are trying to improve the care we offer to women with ovarian cysts by reducing the need for operations to remove them.

What is the problem?

Today you have had a scan, which showed a cyst arising from one of your ovaries. Sometimes it can be difficult to decide whether cysts like yours need to be removed surgically or whether it is safe to leave them alone.

What are we trying to do?

We believe that by measuring different biochemical markers in the blood, we may be able to tell which ovarian cysts may cause problems in the future, requiring an operation to remove them and which can be measured without an operation.

What happens if you decide to take part?

These biochemical markers are measure in the blood. We routinely take a blood sample to check a well-established marker called CA125 and we will take an extra blood sample to check the new biochemical markers at the same time.

How will measuring these markers affect your care?

As the value of these biochemical markers has not yet been investigated, these results will not affect your care. However, if they are found to be useful then they may benefit women in the future.

Do you need to do anything else?

No. This study does not need any extra visits to hospital or blood tests.

What if you decide not to take part?

Participation is voluntary and you may decide not to take part or to withdraw at any time without giving a reason. This will not affect the medical care you receive in any way.

What happens now?

If you agree to take part in the research study then we will ask you to sign a consent form. We will give you a letter for your doctor, which will include a contact telephone number in case of emergency.

Please do not hesitate to ask the doctors or nurses after you if you have any questions. Thank you for your help.

Alex Lawrence
Study Investigator

Davor Jurkovic
Consultant Obstetrician and Gynaecologist

Appendix III: GP letter

Early Pregnancy and Gynaecology Ultrasound Unit

Ruskin Wing

King's College Hospital

Denmark Hill

London SE5 9RS

Dear Doctor,

Your patient has consented to take part in our study "The value of biochemical markers and ultrasound in the management of ovarian cysts". This involves the routine ultrasound scan and CA125 estimation. One extra blood sample has been taken at the time of CA125 estimation for the analysis of novel tumour markers. The study does not involve any other procedures or hospital visits. As the value of these markers has not yet been evaluated, the care of your patient will not be affected. However, if they are found to be of value, they may benefit women in the future.

If you have any questions regarding the study, please contact me on 0171 346 3168.

Yours sincerely,

Mr. Davor Jurkovic

Consultant Obstetrician and Gynaecologist

Appendix IV: Consent form: Study of the value of biochemical markers and ultrasound in the management of ovarian cysts.

I understand the scan I have had today has shown a cyst arising from my ovary. This research study is investigating whether ultrasound scans can be used with markers in the blood to tell whether an ovarian cyst requires surgical removal.

I have been fully informed and have received the patient information leaflet "Study of the value of biochemical markers and ultrasound in the management of ovarian cysts".

All information gained will be confidential and there is no possibility of my being identified.

I agree to have a second sample of blood taken for the measurement of new markers. I understand that taking part in the study will not help my care but may benefit women in the future.

I understand that if I decide not to take part in the study or later withdraw, I do not have to give a reason for my decision and it will not prevent me from obtaining the health care I need.

I have had all my questions answered satisfactorily and I consent to enter this study.

Name.....

Hospital Number .....

Signature of patient .....

Date .....

Signature of investigator .....

Date approved by ethics committee 15/2/2002 Number 00-048

Appendix V: Tumour marker methods

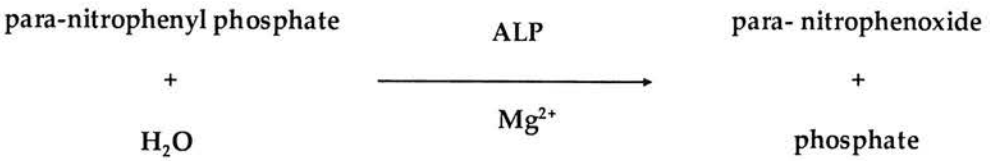
All assays were performed in duplicate and any samples where either of the duplicate values did not fall into within 10% of the other were reassayed. When a quality control value differed by greater than 10% from previous assay means, the assay was repeated.

HER-2/neu

HER-2/neu assay was performed using a commercially available magnetic particle separation immunoassay on the random access automated Bayer Immuno 1 immunochemistry analyzer (Bayer Corporation, New York, USA). The assay uses two monoclonal antibodies (Mab) that bind to independent binding sites on the HER-2/neu extracellular ligand-binding domain (ECD). Reagent 1 contains the Mab NB-3, which is conjugated to fluorescein. Reagent 2 contains a monoclonal Fab fragment, which is conjugated to alkaline phosphatase. The serum samples were thawed at room temperature and mixed thoroughly before use. Reagent 1 (65 μL), Reagent 2 (65 μL), and serum sample or calibrator (20 μL) are incubated for 20 minutes at 37^{0}C. Magnetic particles covalently coated with anti-fluorescein Mab (20 μL) are then added to capture sandwich immunocomplexes. After 28 minutes, the magnetic particles are washed and a colorimetric substrate reagent containing para-nitrophenyl phosphate is added causing an indicator reaction. The alkaline phosphatase (ALP) in the antibody conjugate reacts with the para-nitrophenyl phosphate to form para-nitrophenoxide and phosphate (Figure A.1). The concentration of the HER-2/neu is proportional

to the concentration of para-nitrophenoxide. This is measured photometrically by the rate of increase of light absorbance at a wavelength of either 405 or 450nm.

Figure A.1: The colorimetric reaction of *p*-nitrophenyl phosphate



Evaluation of results

The average absorbency (Y) of each reference standard is plotted against its corresponding concentration (X) to form a calibration curve using a cubic-through-zero curve-fitting algorithm. The value of each sample is determined by simple interpolation from this curve.

Specificity, Sensitivity and Precision

The assay was tested for specificity by the manufacturer, and found to have cross-reactivity with human epidermal growth factor of less than 0.06%. The assay sensitivity was found to have a minimum detectable concentration of 0.1ng/mL. The intra-assay precision was 1.6% and inter-assay precision ranged between 1.1 to 1.7% (Payne, 2000)

Cancer Antigens (CA 15-3, CA 19-9, CA 72-4)

The assay for the cancer antigen 15-3 (CA 15-3) used a commercially available solid phase sandwich immunoassay format (Bayer Corporation, New York, USA). A monoclonal antibody specific to CA 15-3 (115D8 Antibody Conjugate) or Reagent 1 was mixed with the serum sample or control and monoclonal ImmunoMagnetic Particles (*mIMP* ®) and incubated on the random access automated Bayer Immuno 1 immunochemistry analyzer. The DF3 Enzyme Conjugate or Reagent 2 is then added and incubated for a second time during which the antibody complex is bound. Following this, the complex is washed and the para-nitrophenyl phosphate (*p*NPP) substrate is added. The alkaline phosphate in the antibody conjugate reacts with the *p*NPP to form para-nitrophenoxide and phosphate. Increasing absorbance, due to the formation of para-nitrophenoxide, is monitored at 405 nm and 450 nm. Samples with values above the upper range of the calibration curve were diluted with Bayer SETpoint CA 15-3 Assay Zero Calibrator to bring the concentration within the calibration curve and reassayed. A sample with no CA 15-3 will have the minimum label bound, while a sample having a high CA 15-3 will have maximum label bound. Thus, the dose-response curve is proportional to the DF3 reactive determinants in the sample.

A similar method is used for the CA 19-9 assay with the monoclonal antibodies: 1116-NS-19-9 Antibody Conjugate (Reagent 1) and 1116-NS-19-9 Enzyme Conjugate (Reagent 2).

The CA 72-4 assay uses two monoclonal antibodies specific to CA 72-4: cc49 Antibody Conjugate (Reagent 1) and B72.3 Enzyme Conjugate (Reagent 2).

The reagents for all assays are light sensitive so once opened were stored in the dark at four degrees centigrade.

Evaluation of results

The average absorbency (Y) of each reference standard is plotted against the corresponding concentration (X) to form a calibration curve using a cubic-through-zero curve-fitting algorithm. The value of each sample is determined by simple interpolation from this curve.

Sensitivity and Precision

The minimum sensitivities (Table A.1) and precision (Table A.2) are shown below.

Table A.1: Minimum sensitivities of cancer antigen assay kits (data supplied by manufacturer with assay kit).

Cancer Antigen	Minimum detectable (U/mL)
15-3	0.2
19-9	0.8
72-4	0.3

Table A.2: The intra and inter-assay coefficients of variation for cancer antigen assays (data supplied by manufacturer with assay kit).

Cancer Antigen	Intra-assay Coefficient of Variation (%)	Inter-assay Coefficient of Variation (%)
15-3	3.0	4.0
19-9	3.5	4.5
72-4	3.0	6.0

Cancer Antigen 125

CA 125 assay was performed as previously described using a heterologous assay using monoclonal antibodies M11 and OC125 incubated on the random access automated Bayer Immuno 1 immunochemistry analyzer (CA 125 II, Bayer Corporation, New York, USA) according to the manufacturer's instructions.

Appendix VI: Conference proceedings

Lawrence AC, Aslam N, Elson CJ, Salim ZR and Jurkovic D. Incidental discovery of small adnexal tumours: how well do the conventional diagnostic models work? 29th British Congress of Obstetrics and Gynaecology. Birmingham, UK. July, 2001.

Lawrence AC, Aslam N, Woelfer B, Elson CJ and Jurkovic D. Diagnosis of ovarian cancer with ultrasound, serum CA125 and logistic regression. 10th World Congress on Ultrasound in Obstetrics and Gynaecology.
Zagreb, Croatia. October, 2000.

Lawrence AC, Aslam N, Elson CJ, Salim ZR and Jurkovic D. Incidental discovery of small adnexal tumours: how well do the published models work? 11th World Congress on Ultrasound in Obstetrics and Gynaecology. Melbourne, Australia. October, 2001.